Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 22 (2020), Issue 12 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Tutorial

Original Papers

Corrigenda and Addendas

Review

Involvement of People With Dementia in the Development of Technology-Based Interventions: Narrative Synthesis Review and Best Practice Guidelines

Harleen Kaur Rai[1], BSc, MSc; Aline Cavalcanti Barroso[1,2], BSc, MSc; Lauren Yates[3], PhD; Justine Schneider[2], PhD, Prof Dr; Martin Orrell[1], PhD, Prof Dr, FRCPsych

[1]Division of Psychiatry and Applied Psychology, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom
[2]School of Sociology and Social Policy, Law and Social Sciences Building, University of Nottingham, Nottingham, United Kingdom
[3]Department of Clinical, Education and Health Psychology, University College London, London, United Kingdom

Corresponding Author:
Harleen Kaur Rai, BSc, MSc
Division of Psychiatry and Applied Psychology
Institute of Mental Health
University of Nottingham
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 01157484252
Email: Harleen.Rai@nottingham.ac.uk

Abstract

Background: Technology can be helpful in supporting people with dementia in their daily lives. However, people with dementia are often not fully involved in the development process of new technology. This lack of involvement of people with dementia in developing technology-based interventions can lead to the implementation of faulty and less suitable technology.

Objective: This systematic review aims to evaluate current approaches and create best practice guidelines for involving people with dementia in developing technology-based interventions.

Methods: A systematic search was conducted in January 2019 in the following databases: EMBASE (Excerpta Medica database), PsycINFO, MEDLINE (Medical Literature Analysis and Retrieval System Online), CINAHL (Cumulated Index to Nursing and Allied Health Literature), and Web of Science. The search strategy included search terms in 3 categories: *dementia*, *technology*, and *involvement in development*. Narrative synthesis wove the evidence together in a structured approach.

Results: A total of 21 studies met the inclusion criteria. Most studies involved people with dementia in a single phase, such as development (n=10), feasibility and piloting (n=7), or evaluation (n=1). Only 3 studies described involvement in multiple phases. Frequently used methods for assessing involvement included focus groups, interviews, observations, and user tests.

Conclusions: Most studies concluded that it was both necessary and feasible to involve people with dementia, which can be optimized by having the right prerequisites in place, ensuring that technology meets standards of reliability and stability, and providing a positive research experience for participants. Best practice guidelines for the involvement of people with dementia in developing technology-based interventions are described.

(J Med Internet Res 2020;22(12):e17531) doi:10.2196/17531

KEYWORDS

dementia; technology; co-production; participation; development

Introduction

Background

Technology can be used to address some of the challenges of dementia care and enable people with dementia to maintain their independence for as long as possible [1]. Despite the wide variety of technology available (eg, reminder devices, touchscreen devices and apps, and computerized cognitive and physical interventions) [2], there is a lack of evidence on their efficacy, and many interventions are either in the development phase or in a prototype phase [3]. Moreover, there has been little involvement of people with dementia in the development of technology-based interventions [4]. Possible reasons for this lack of involvement include stigma, concerns about the frailty of older people, and the anticipated distress among participants caused by trying out less developed information technology [5]. Underdeveloped technology-based interventions with inadequate involvement could have residual faults and could potentially make early prototypes harder to operate for people with dementia and lead to people with dementia being reluctant to use them [5]. Consequently, technologies are being developed that are neither user-friendly nor fit for the purpose of supporting people with dementia [3,4]. Technology that is faulty or poorly designed may not be useful in supporting people with dementia.

A previous systematic review showed that people with dementia are able to provide useful feedback, such as comments on screen size, language difficulties, and the importance of personalization on private spaces of websites, which help to improve the quality of the intervention [4]. This approach improves the usability and acceptability of the technology-based interventions [4] and generates enjoyment and enthusiasm in participants with dementia [6,7]. However, Span et al [4] only reviewed papers up to 2010, and subsequently, many innovations in technology have taken place. Furthermore, Astell et al [8] and Span et al [4] assert that to optimize technology by ensuring the needs and preferences of people with dementia are addressed, it is crucial to implement a participatory process in which people with dementia are involved throughout the development process [4,8].

Information on how to optimize the involvement of people with dementia is dispersed, and there is a clear need to bring the evidence together in a systematic manner through an appraisal of the involvement of people with dementia in the development of technology-based interventions and guidelines on how to best facilitate and optimize this involvement.

Objectives

This narrative synthesis systematic review sets out to appraise the methods used by applying existing frameworks, such as the Medical Research Council (MRC) framework for the evaluation of complex interventions and the Centre for eHealth Research (CeHRes) roadmap [9,10], and to create best practice guidelines on how to better involve people with dementia in developing technology-based interventions accompanied by a logic model.

Methods

Narrative Synthesis

Narrative synthesis is “an approach to the systematic review and synthesis of findings from multiple studies that relies primarily on the use of words and text to summarize and explain the findings” [11]. Narrative synthesis can be used to address a multitude of questions regarding the effectiveness of interventions, including what works but also why and how. Narrative synthesis is preferred for this review as it can be used to convert the evidence into clear and structured best practice guidelines on how to facilitate the participation of people with dementia in the development of technology-based interventions. The approach consists of 4 elements: theory development, developing a preliminary synthesis, exploring relationships within and between studies, and assessing the robustness of the synthesis.

Element 1: Theory Development

Theory development underpins the systematic review by supporting the development of the review question and the types of studies to be included. Our starting point is the desirability of end user involvement in technology development. Several studies suggest that feedback from people with dementia can lead to improvements in the overall quality of the technology [4,8]. This would result in more useful and suitable pieces of technology and would also increase the willingness to use the technology. Furthermore, the involvement of end users in developing technology could also support the implementation of a technology in the future, leading to a better range of technology to improve the quality of life of people with dementia. Therefore, we only include studies that clearly illustrate how feedback was gathered from people with dementia during development. This would exclude studies with a sole focus of including participants as objects of studies where no meaningful involvement has taken place. The narrative synthesis undertaken here will contribute to the refinement of our theoretical starting point and support the application of the findings of this review [11].

Element 2: Developing a Preliminary Synthesis

The preliminary synthesis develops an initial description of the results of the included studies, organized in a manner such that a pattern can be described in terms of effects or impact [11]. This can be done through the use of textual descriptions, grouping and clusters, and tabulation. This preliminary synthesis is necessary to inform the next steps of the narrative synthesis.

Element 3: Exploring Relationships Within and Between Studies

The patterns that emerge from the preliminary synthesis are subjected to a more detailed analysis in which the reviewers move toward exploring the relationships within and across the included studies [11]. The relationships between the characteristics and reported findings of different studies are reviewed. This element of narrative synthesis will help identify the factors that may have influenced the results and will seek to provide an explanation of how and why a particular intervention works [11]. The methods used here include qualitative case descriptions and the development of a conceptual model based on the grouping of study findings. This will help to structure the inferences drawn from our results.

Element 4: Assessing the Robustness of the Synthesis

The final element of narrative synthesis sets out to review the trustworthiness of the results [11]. The trustworthiness of the synthesis is affected by the quality and quantity of the evidence on which the synthesis is built and by the methods used. Therefore, an appraisal is undertaken to judge the strength of the evidence for the findings and to generalize them to different populations and contexts [11].

Electronic Searches and Screening

This review was registered in the International Prospective Register of Systematic Reviews (PROSPERO) under protocol number CRD42017068933. After conducting 2 pilot searches, we systematically searched the following databases: EMBASE (Excerpta Medica database), PsycINFO, MEDLINE (Medical Literature Analysis and Retrieval System Online), CINAHL (Cumulated Index to Nursing and Allied Health Literature), and Web of Science in January 2019. Studies published between 2000 and 2019 were considered. The search strategy consisted of combinations and variations of search terms in the following 3 key categories: *dementia*, *technology*, and *involvement in development*. Involvement terms also included *codesign*, *participatory research*, and *user participatory development*.

After removal of the duplicates, a 3-stage screening process was independently conducted by 2 review team members (HR and AB): (1) titles were screened for relevance to the review question, and irrelevant studies were archived; (2) abstracts were assessed (referring to the full text whenever necessary to clarify the relevance of the study); and (3) quality assessment of the remaining studies was conducted (see the *Data Extraction and Study Quality Assessment* section). The reasons for exclusion were recorded by archiving the excluded studies in relevant folders in EndNote (Clarivate Analytics). In case of disagreement between the 2 reviewers, a third review team member was consulted (LY). Additional studies from the review by Span et al [4] were distributed separately among 4 review members (AB, JS, HR, and LY) for data extraction and quality assessment. The reference lists of studies that passed the quality assessment were reviewed to ensure the inclusion of other relevant papers.

Criteria for Inclusion and Exclusion of Studies

The inclusion and exclusion criteria of the studies were as follows:

- Types of participants: people with a diagnosis of dementia, irrespective of age, type of dementia, or stage of the disease.
- Types of intervention: involvement of people with dementia in the development process of a technology-based intervention.
- Types of studies: quantitative, qualitative, and mixed methods studies published from the year 2000 onward as an English language journal paper with sufficient study quality (a minimum of 5 criteria met as assessed with the Critical Appraisal Skills Programme [CASP] guidelines or 50% of the criteria met as assessed with the Downs and Black checklist).

Description of Development Phases

The development process of a technology-based intervention consists of several stages. To identify the key stages of technology development for this review, we employed the MRC framework together with the CeHRes roadmap [9,10]. Both frameworks focus on developing interventions; however, although the MRC framework is more widely used for developing complex interventions, the CeHRes roadmap has a focus on digital health interventions (Table 1).

Table 1. Description of the Medical Research Council framework and the Centre for eHealth Research roadmap.

Phase	Medical Research Council framework	Centre for eHealth Research roadmap
Development	• Single phase • Identifying evidence base (eg, systematic review) • Identifying or developing theory (eg, scope existing theories and interviewing stakeholders) • Modeling process and outcomes (eg, undertaking a pretrial economic evaluation, focus groups, surveys, and case studies)	• Multiple phases such as contextual inquiry, value specification, and design • Identifying problems and needs of intended users (eg, literature review, field observations, interviews, and workshops) • Determining the most favorable solutions based on the values of the stakeholders • Building prototypes to fit values and user requirements (eg, focus groups and field testing)
Feasibility and piloting	• Specific phase for feasibility and piloting • Activities consist of testing procedures for acceptability, determining appropriate sample size, and estimating rates of recruitment	• N/A[a] (can be part of the design phase)
Evaluation	• Assessing clinical and cost-effectiveness (eg, randomized controlled trial) • Understanding processes (process evaluation)	• Summative evaluation • Assessment of the impact of eHealth technologies in clinical, organizational, and behavioral terms
Implementation	• Getting evidence into practice • Surveillance, monitoring, and long-term outcomes	• Operationalization • Activities to introduce, adopt, and employ the technology in practice (eg, creating a business model)

[a] N/A: not applicable.

Data Extraction and Study Quality Assessment

A standardized data extraction form was developed by the primary researcher (HR), in which the review team members recorded the extracted data from the final studies, including the study quality rating (Multimedia Appendix 1) [6,7,12-30].

Quality was assessed using the CASP guidelines. These guidelines consist of 8 checklists for various types of studies and include items that assess multiple aspects of research (eg, recruitment, risk of bias, confounders, data collection, data analysis, results, and implications) [31]. The studies were rated as high quality if 8 or more criteria were met, medium quality if 5 to 7 criteria were met, and low quality if 4 or less criteria were met [32]. Studies that did not meet the criteria for assessment with the CASP guidelines were assessed with the Downs and Black checklist [33]. This checklist is appropriate for both randomized and nonrandomized studies and consists of 27 items over 5 domains (reporting, external validity, internal validity—bias, internal validity—confounding, and power). The maximum score was dependent on the study design; however, each study was rated as high quality if it met over 81% of the criteria, medium quality if it met over 66% to 80% of the criteria, fair quality if it met over 51% to 65% of the criteria, and low quality if it met 50% of the criteria or less [34]. Studies considered to have low quality were excluded. The review team members independently assessed the studies for sufficient study quality. Any differences in judgment between the 2 reviewers were resolved by a third review team member.

Consultation With the Patient and Public Involvement Group

One reviewer (HR) presented the findings at a patient and public involvement (PPI) consultation meeting on 2 different occasions. This PPI group is run on a monthly basis at the Institute of Mental Health in Nottingham. The aim of both meetings was to gain insights into people’s own views on optimal involvement in developing technology-based interventions; their feedback and comments on the findings; and, more specifically, their feedback on the guidelines drafted by the authors. This feedback would then be integrated within the findings of this review and used to strengthen the best practice guidelines.

The first meeting was attended by 2 people with dementia, 1 carer, 1 volunteer, and 1 researcher and lasted for 45 min. The second meeting was attended by 2 people with dementia, 2 carers, 1 volunteer, and 4 researchers and lasted for 25 min. After a brief introduction to the review and its findings, the best practice guidelines were presented one at a time on a projector. In the first meeting, printed handouts were distributed to each participant. A short discussion in terms of relevance and accuracy encompassed each guideline, and notes were taken throughout the meeting.

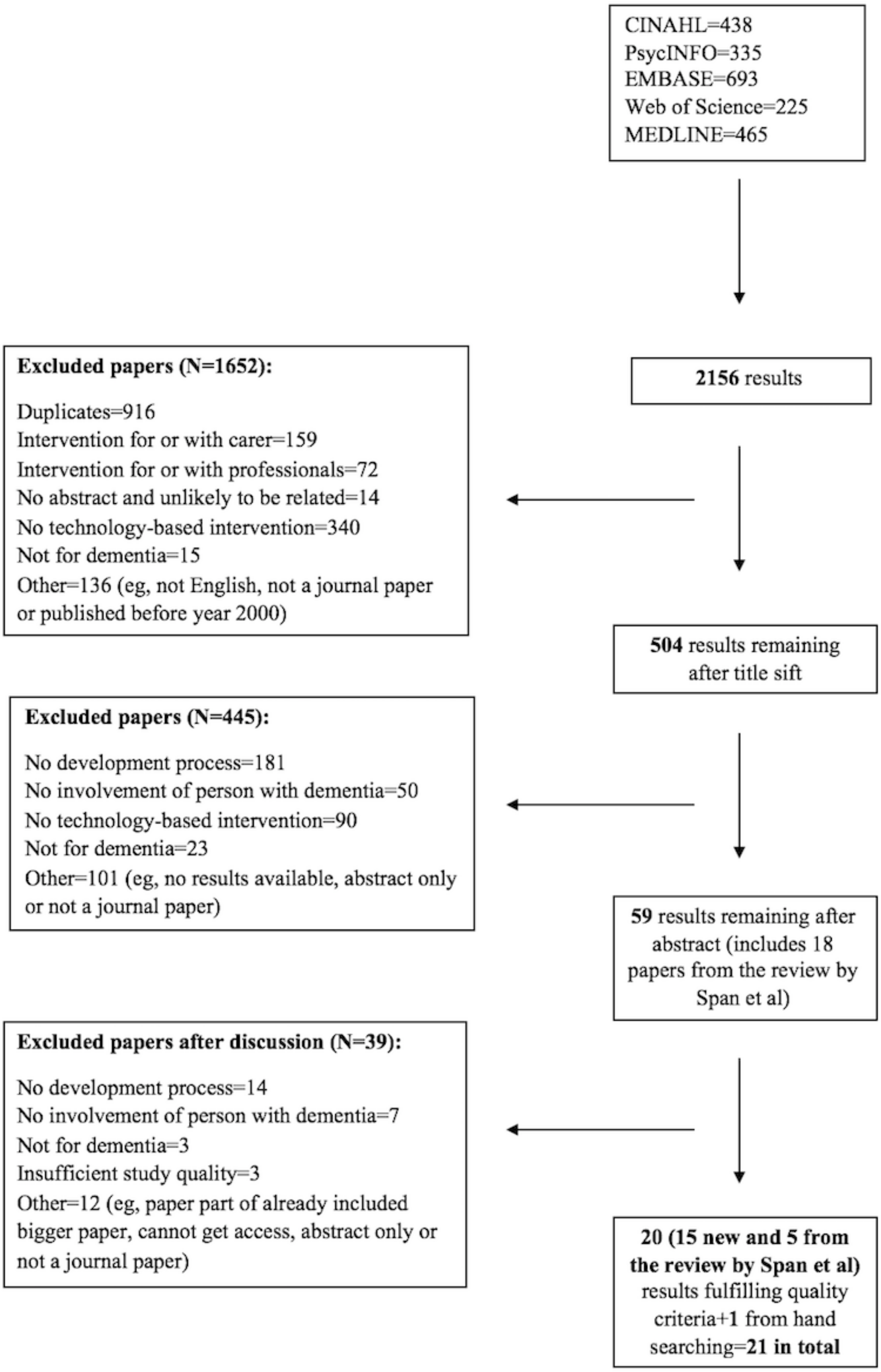
Results

The *Results* section comprises the second element of narrative synthesis: developing a preliminary synthesis.

Search Results

A total of 2156 potentially relevant titles were identified across the 5 databases (Figure 1). Removal of duplicates and screening of titles, abstracts, and full texts resulted in 20 studies that met the inclusion criteria. The most frequent reasons for exclusion were the lack of a technology-based intervention and absence of a development process. Additional hand searching led to the inclusion of one other study, making up a total of 21 studies. This study came from a review by Span et al [4], which was not captured by the current search strategy. Other studies from the same review not captured by the search strategy (n=7) were excluded because they did not meet the inclusion criteria (eg, not a journal paper or low study quality). The reference lists of studies passing the quality assessment were reviewed to ensure that any other relevant studies would be included.

Figure 1. Flowchart of study selection. CINAHL: Cumulated Index to Nursing and Allied Health Literature; EMBASE: Excerpta Medica database; MEDLINE: Medical Literature Analysis and Retrieval System Online.



Description of Included Studies

The main study characteristics of all 21 studies included study sample and design, description of the technology-based intervention, and rating of study quality (Multimedia Appendices 1 and 2) [6,7,12-30]. Using the CASP Qualitative checklist, 11 studies were assessed as high quality and 8 studies were assessed as medium quality. Only 1 study was assessed with the CASP Randomized Controlled Trial (RCT) checklist, which met 7 of 11 criteria [12]. Another study was assessed using the Downs and Black checklist. It was rated as fair quality, meeting 65% of the criteria for a before and after follow-up study [13]. Most studies were conducted in Europe (n=17), 3 studies took place in Australia [13-15], and 1 was conducted in Canada [16].

A majority of the studies adopted a purely qualitative methodology (n=14). A total of 6 studies employed a mixed methods approach, of which 1 combined qualitative methods with a controlled trial [12]. Only 1 study adopted a purely quantitative methodology [13]. The studies described a variety of technology-based interventions, including communication aids, music tools, devices to support activities of daily living, reminder systems, and tracking devices. In the majority of the studies, people with dementia were involved along with carers or other professionals who either supported the person with dementia in their involvement or provided separate input themselves (n=17). Only 4 studies solely included people with dementia [13,17-19].

Methods of Involvement and Key Findings

The methods used to involve people with dementia along with the phases of the MRC framework and CeHRes roadmap are summarized in Textbox 1, allowing for an initial synthesis of the findings.

Textbox 1. Methods used to involve people with dementia in the studies (N=21) according to the Medical Research Council framework phases.

Development (contextual inquiry, value specification, and design)

- Behavioral observations [17,20], focus groups [6,7,18,21-24], interviews [6,18,19,22,24-27], workshops [7,25], questionnaires [17], and user tests [6,7,18-20,22]

Feasibility and piloting

- Behavioral observations [13-16,28,29], interviews [14-16,25,27,29,30], questionnaires [13,16,29,30], field testing [18,25,29], and technical system usage [27]

Evaluation (summative evaluation)

- Randomized controlled trial [12], focus groups [12], interviews [12], and questionnaires [12]

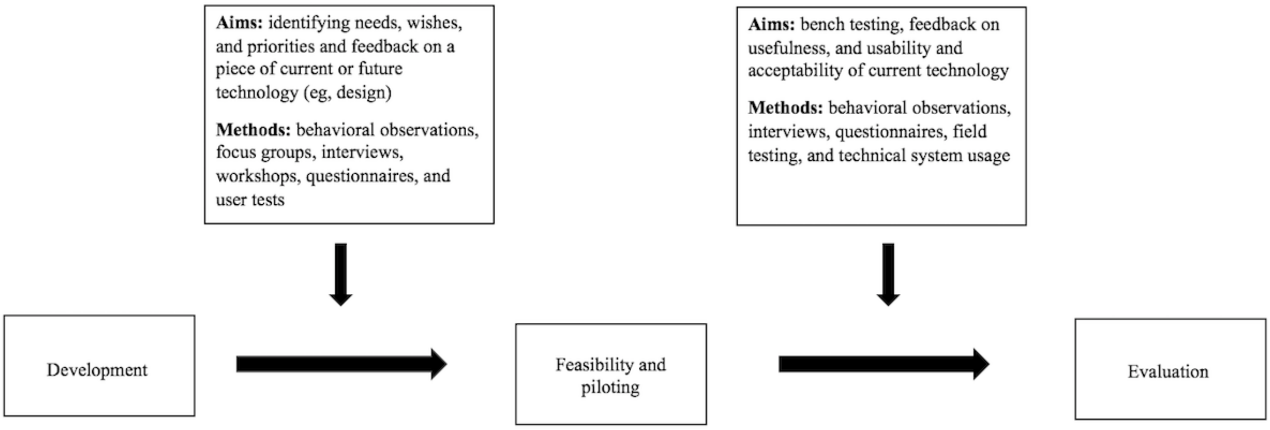
Implementation (operationalization)

- Not applicable

Development Phase (n=10)

A total of 10 studies involved people with dementia solely in the development phase, which coincides with the contextual inquiry, value specification, and design phase of the CeHRes roadmap. The majority of these studies primarily employed qualitative methods, such as focus groups and semistructured interviews. At times, these were accompanied by user tests, observations, and questionnaires. Textbox 1 provides an overview of all methods used in the development phase. The aims of the studies ranged from identifying the needs, wishes, and thoughts of people regarding certain areas for development (eg, independence or cognitive reinforcement) to gaining feedback on the design of future or existing technologies (Figure 2).

Figure 2. Aims and methods of involvement along the development stages of technology according to the Medical Research Council framework.



Needs Assessment and Design of Future Technology

A total of 2 studies included needs assessments, followed by discussions about the design of future technology using qualitative methods. Boman et al [21] used focus groups to capture experiences, expectations, and thoughts concerning a videophone and its design concepts. The design had to be flexible to meet the needs of people with dementia, be easy to use, and not look like assistive technology. Another example is the study by Robinson et al [7] who also used focus groups to elicit views and concerns about independence from people with dementia and carers. A list of priorities was derived from the findings. Areas for functional improvement included 2-way communication, flexibility of functionality, and something to *guide* them home when outside. Workshops were then used to identify the preferred design and functionality aspects of future technologies. Finally, user tests were performed with paper prototypes until 2 fully functional devices were developed.

Needs Assessment and User Tests

In 3 other studies, needs assessments were followed by user tests with functional technologies. Orpwood et al [19] used interviews with users (user surveys) to compile a wish list of issues that were of importance in maintaining the quality of life of people with dementia. A large list of potential technologies that could address these issues was generated. The following 4 technologies were selected for initial development: a music player, a device to reduce social isolation, a conversation prompter, and a device to support sequences of tasks. Useful design guidelines were derived from user tests, particularly for intuitive control interfaces (eg, controls need to stand out, be big, and simple).

Touch screens appear to be very intuitive, and prompts seem to be more effective than verbal or visual instructions. Hanson et al [6] used focus groups to identify user needs and preferences and to structure the material within a multimedia program. A prototype was taken forward in user tests followed by in-depth interviews. These led to the identification of problems such as logging into and out of the program and accessing the exercises. Participants enjoyed the computer training sessions and gained considerable satisfaction from learning a new skill that they previously thought was not feasible. Lopes et al [22] used interviews to analyze user needs and identify commonly misplaced items, such as keys, glasses, cell phones, and identity papers. Focus groups and user tests were then used to try out existing item locators and define the following system requirements of a new item locator prototype: ease of use, capacity for customization, low price, nonstigmatizing design, and being *fun* to use. The next step included user tests with the first prototype in which participants commented that they would prefer to be guided by a customized sound of a voice system to find an item.

Design of Existing or Future Technology

In 2 studies, feedback was gathered solely on the design of future technologies using qualitative methods only. In a study by Meiland et al [24], nonfunctional mockups were reviewed after discussing the potential functionalities of an integrated, assistive system in focus groups and interviews. Participants valued help in case of emergencies, navigation support, and the calendar function the most. The least preferred functionalities were activity support and picture phone dialing. McCabe and Innes [23] found that people with dementia and carers gave specific feedback on the form and features of a potential GPS design during focus groups (eg, waterproof watch style design with a range of colors). However, participants would have preferred to comment on an actual and active device rather than talking hypothetically, as it did not provide them with enough context.

In 3 studies, participants gave feedback on the design of an existing and functional technology. Freeman et al [17] analyzed observational data of people with dementia using 2 websites. These data helped to uncover 3 major problems: scrolling, nonrecognition of more information on a page, and getting stuck. There was a high degree of overall satisfaction with both sites measured through questionnaires. Kerkhof et al [26] interviewed residents after bench testing a memory aid (planning board). The majority of residents appreciated the use and function of the aid, but successful implementation was difficult because of installation errors, limited user friendliness, and lack of knowledge regarding the function and the use of the aid. Areas of focus for improvement include software program adaptation, additional technological applications, internet connectivity, accessibility, and addition of media. Finally, Klein et al [20] also analyzed observational data of the participants while testing 2 prototype devices. On the basis of the findings from these tests, a third prototype device was developed. Special attention was given to more personally relevant and engaging content, contextual factors, higher levels of immersions, and more control for the user.

Feasibility and Piloting Phase (n=7)

A total of 7 studies included only the feasibility and piloting phase, which can be part of the design phase of the CeHRes roadmap [13-16,28-30]. In this phase, people with dementia were given the opportunity to try out a piece of technology in a pilot study or through field testing. Often, the aim was to gain insights into the usefulness of a device along with its acceptability and usability (Figure 2). In the majority of the studies (n=5), a mixed methods approach was adopted where participants were observed while using the device and feedback was obtained through semistructured interviews and questionnaires. Textbox 1 gives an overview of all the methods used in the feasibility and piloting phase.

Begum et al [16] used observations to investigate adherence to prompts from a robot, engagement with the robot, and how often a task was completed. Interviews and questionnaires provided information on the acceptance, ease of use, usefulness, and physical attributes of the assistive robot. Meiland et al [29] field tested an integrated digital prosthetic with multiple functionalities. Data on its usability were collected through behavioral observations, interviews, and questionnaires, and it was deemed to be user-friendly and useful, but there was a desire for more personalization and configuration of reminders.

Moyle et al [15] explored the acceptability of a telepresence robot using observations through video recordings and follow-up interviews. Participants indicated a positive social presence, which was also observed through the display of positive emotions. A similar methodology was adopted in another study by Moyle and Jones [14]. Observations through video recordings were used to describe the effectiveness of a virtual reality forest (VRF) on engagement, apathy, and mood states. Overall, the VRF was perceived to have a positive effect, but there were higher levels of fear and anxiety. Follow-up interviews were used to explore the experiences of using the VRF. Most participants reported positive perceptions and suggested making the experience more active.

Topo et al [30] used questionnaires to collect information on the functional ability of people with dementia. Through interviews, data were collected on the usage and usefulness of an existing music tool 2 weeks after installation in a care home. Most participants benefited from its use and had positive experiences. Some problems were reported with the sensitivity of the touch screen and the font size being limited because of the screen size.

Jamin et al [28] used a qualitative approach where participants were involved in usability testing and were observed while interacting with *VENSTER*. The content of VENSTER, which needs to provide enough context to be meaningful, was interesting and suitable for the participants. The study by Khosla et al [13] was the only study using a quantitative methodology in which participants were observed while interacting with a social robot to gain insights into emotional, visual, and behavioral engagement. In addition, user surveys were used to assess acceptability. The participants generally had a positive attitude toward social robots. Most of the participants gave high ratings in terms of the perceived usefulness and enjoyment of their experience with the robot.

Development and Feasibility and Piloting Phase (n=3)

A total of 3 studies elaborated on both the development, and feasibility and piloting phase [18,25,27]. These studies systematically described the involvement of people with dementia over the course of each phase: the identification of user needs and wishes, determination of the design, and testing of a prototype version through a pilot test or field test. For each of these activities, a wide array of methods was applied, such as focus groups and interviews, workshops, and usability tests.

In the study by Span et al [18], the development phase consisted of interviews to identify needs and preferences for an interactive web tool and focus groups to discuss the results of the interviews and to make any additions to the problems and experiences shared. Several user requirements were identified, such as social contacts, daily activities, care, autonomy, involvement, and communication, specifically for the decision-making process. Paper mockups were discussed in focus groups to design the interactive prototype. Thereafter, individual user tests were organized to gather feedback on an interactive prototype regarding design, content, and user friendliness. Some participants found it difficult to comment on paper mockups but overall mentioned that information per screen and the number of screens should be decreased and the accuracy of language was of importance. For the feasibility and piloting phase, an interactive prototype was field tested to gain feedback on the user friendliness of the tool, the contentment of the participants, and how they valued the tool for decision making.

Martin et al [27] used interviews in the development phase to identify the main issues, risks and care needs arising during nighttime. The main themes included promoting independence, maintaining dignity, maximizing social inclusion, managing risk, and providing stimulation. In the feasibility and piloting phase, participants were involved in any of the 3 phases of iterative validation and evaluation of a prototype through technical system usage and interviews. The phases included testing for stability, usability, and integration within a full telecare system and the implementation of music and light. Participants liked the mobile component of the nighttime system and the easy navigation.

Davies et al [25] used both interviews and workshops to identify user needs in specific areas of cognitive reinforcement in the development phase. The following areas were identified by the participants: remembering, maintaining social contact, performing daily life activities, and enhanced feelings of safety. Interviews accompanied field testing in the feasibility and piloting phase. After trying out 4 prototypes, participants highlighted the need for personalization, less complex functionality, and extended use within the home environment.

Evaluation Phase (n=1)

One study involved people through evaluation in a controlled trial [12]. Participants used an assistive system and filled in posttest questionnaires to assess impact. Despite no significant effects on impact, posttrial interviews and focus groups were used to assess qualitative impact, and participants found the system to be very useful but not user-friendly because of technical difficulties, including the unresponsiveness of touch screens and issues with gaining access. For people who had not used a touch screen before, the system was deemed unintuitive.

Involving People With Dementia

Impact on the Developed Technology

In all but 5 studies [13,15,19,20,25], researchers directly reflected on the involvement of people with dementia in the development of the technology-based intervention. Researchers concluded that it was both necessary and feasible to involve people with dementia throughout the development process. In addition, Kerkhof et al [26] argued that it is not sufficient to respond to the needs of people with dementia by solely involving carers or staff members. This is further supported by Meiland et al [24] and Lopes et al [22] who found that exploring the user perspectives from various stakeholders, including people with dementia, is necessary to understand the problem and come up with possible solutions. Jamin et al [28] also emphasized that codesign with all stakeholders can make the overall experience more pleasurable but also more meaningful, as it allows the users to be kept at the center of the decision-making process and adaptations can be made to new insights as they emerge. In several studies, it was recognized that people with dementia continue to be one of the most excluded groups from research and the design of new services [6,21]. Possible reasons for this could be difficulties in recruitment or the cognitive impairment of people with dementia [16,29]. However, despite these challenges, all studies recommended the involvement of people with dementia in future studies, as this could lead to obtaining views on new concepts or ideas for technology and to more concrete feedback on the usability and user friendliness of a device. For instance, one study determined how to maximize website suitability for people with dementia after receiving feedback [17]. Another study adapted the appearance of a robot and made it more socially interactive [16]. Finally, people with dementia suggested that the interaction between end users and a virtual reality system could be improved by incorporating reminiscence within the tool [14].

Impact on the Person With Dementia

The positive effects of involvement on people with dementia themselves included the empowering effects of involvement that were evident in increased feelings of well-being, being able to voice opinions, learning a new skill through the use of technology, and an enhanced sense of control experienced by the majority of the participants [6]. Participants were also motivated to contribute to research and a better quality of life for future people with dementia [18,26]. No distress or adverse events from involving people with dementia were reported in any of the studies.

Outcomes of the PPI Consultations

PPI group members reflected on how to optimize involvement both in research and in the development of technology-based interventions and endorsed the guidelines (Multimedia Appendix 3) [6,12,18,21,23,26-28,30]. Additions were made to some guidelines, for example, there was consensus among members that researchers need to focus on individual research participants, which includes awareness of their type of dementia, any other relevant conditions, and any specialized knowledge of participants, which could further support the development of technology. Awareness among participants in terms of the relevance and positive effects of involvement for them was also important.

A friendly research environment was helpful to make people feel comfortable to ask questions in case they did not understand something. This is especially helpful when developing new technology, which can include some unknown aspects, and so researchers should also aim to avoid abbreviations and acronyms to avoid technology-related jargon. In addition, PPI group members suggested that researchers should present their materials at a PPI meeting before an actual research activity takes place to ensure the use of jargon is limited.

Involving people with dementia as early as possible in the development process and in multiple phases of development should lead to increased familiarity and a better understanding of the technology. Members were also positive about encouraging technology developers to interact directly with people with dementia but highlighted that a mediator (eg, a researcher) would be necessary to ensure a good level of understanding between people with dementia and the developers. A person with dementia also mentioned taking a technology into the community (eg, a memory café) to gather feedback, as this would allow for the technology to be used in a real-life setting.

Finally, a *Wizard of Oz* method was suggested by a researcher where participants interact with a working prototype but under the guidance of an unseen researcher. The 2 PPI group members with dementia mentioned that they would not have an issue with this in terms of ethics, and it was regarded as a good idea. This method could be used to limit the amount of errors.

Discussion

The *Discussion* section comprises the third element of narrative synthesis: exploring relationships within and between studies.

Summary and Interpretation of Findings

People with dementia can contribute effectively to the development of technology but are often excluded from research in this area. With the rise of innovative technology, there is a need for an overview of the current evidence regarding the involvement of people with dementia and recommendations on how to optimize this involvement in the development process. This is to ensure that the developed technologies are suitable and tailored toward the needs of the end users. This is the first narrative synthesis review to synthesize the findings from high-quality studies of involvement of people with dementia in developing technology-based interventions and has created best practice guidelines based on the evidence summarized below.

One of the strengths of this review is the strict inclusion criteria leading to the synthesis of high-quality papers. This further supports the robustness of the findings and the developed guidelines. Furthermore, the application of narrative synthesis in this systematic review allowed for a highly systematic approach to search for and make sense of the evidence. The underpinning theory, as part of the first element of narrative synthesis, helped define the research questions and the studies to be included in this review. In addition, the preliminary synthesis supported the tabulation of the findings, which is highlighted in the text, tables, and figures. This approach also proved helpful in converting the evidence into best practice guidelines by looking for relationships within and between the studies. Good examples of involvement were extracted and incorporated into the guidelines, which were modified by the input of the PPI group. This enabled the invaluable perspective of people with lived experience on the findings and helped strengthen the robustness of the synthesis and relevance of the guidelines.

The findings suggest that the involvement of people with dementia varies depending on the development stage and methods used, which is in line with previous research [4]. A large part of involving people with dementia revolves around identifying user needs and preferences. The majority of the studies included this aspect in their research and primarily used qualitative methods such as focus groups and interviews. The identification of needs often helped to prioritize the most pressing issues for people with dementia.

Another component is gathering feedback on either the prospective or existing design of a device. These activities mostly include qualitative methods while using observations and questionnaires. People with dementia play an active role in voicing their opinions and trying out the available prototypes. Once a piece of technology has been developed into a more refined version, the involvement of people with dementia shifts more toward the participants becoming the objects of study. In several studies, people with dementia were asked to use a piece of technology more rigorously during a field-testing phase accompanied by observations and to give feedback after the test phase. Interestingly, no studies involved the participation of people with dementia in the implementation phase.

These findings are congruent with findings from a review by Span et al [4]. However, in this review, we found studies that described more elaborately the involvement of people with dementia and demonstrated that it is feasible to include them throughout the entire development process rather than in a single phase. The involvement of people with dementia started with exploring their needs and gaining an understanding of the current problem, which led to the development and testing of various prototypes together with people with dementia to tailor it to their needs. These studies set a good example for future studies by applying various methods and obtaining in-depth data from people with dementia. The impact of the involvement is also evident as studies have provided examples of concrete pieces of feedback from people with dementia, which improved the developed technology. However, there is also an impact of involvement on the person himself or herself, as some studies have shown that the involvement of people with dementia can be empowering and lead to increased feelings of well-being [6]. Participants expressed the importance of being able to contribute to the research by voicing their own opinions [4,6,26]. None of the studies noted any distress caused by the involvement of people with dementia. This is helpful for future studies, as anticipated distress from trying out underdeveloped technology was seen as a reason to not include people with dementia in development [19].

Some challenges were described in the involvement of people with dementia, such as the risk of obtaining socially desirable answers [21,29]. However, this risk is not specific to this population and, in general, is not uncommon in research. Another challenge was obtaining in-depth feedback from participants, as the use of unfamiliar terms related to technology made it difficult for participants to comprehend the questions [30]. Jamin et al [28] emphasized the need for the involvement of multiple stakeholders but acknowledged that this adds a level of complexity to the design process, as researchers or developers would have to navigate various differing opinions. Despite these challenges, all studies recommended that people with dementia should be involved in developing technology and also to keep including relevant stakeholders such as (in)formal carers and technology developers where possible.

Best Practice Guidelines (Narrative Synthesis Element 4: Assessing the Robustness of the Synthesis)

On the basis of the findings from the studies included in this review and the contributions from the PPI consultation meetings, best practice guidelines for the involvement of people with dementia in developing technology-based interventions were developed (Multimedia Appendix 3). A previous best practice model included in a systematic review by Di Lorito et al [35] served as an example to better organize the findings according to the goals of involvement, preparations, and the contributions from the PPI consultation meetings. A score can be allocated to each guideline depending on whether it has been fully met (score=2), partly met (score=1), or not met (score=0). The availability of 12 guidelines means that a total score of 24 can be achieved, indicating that each guideline has been met in full when developing a technology-based intervention for people with dementia.

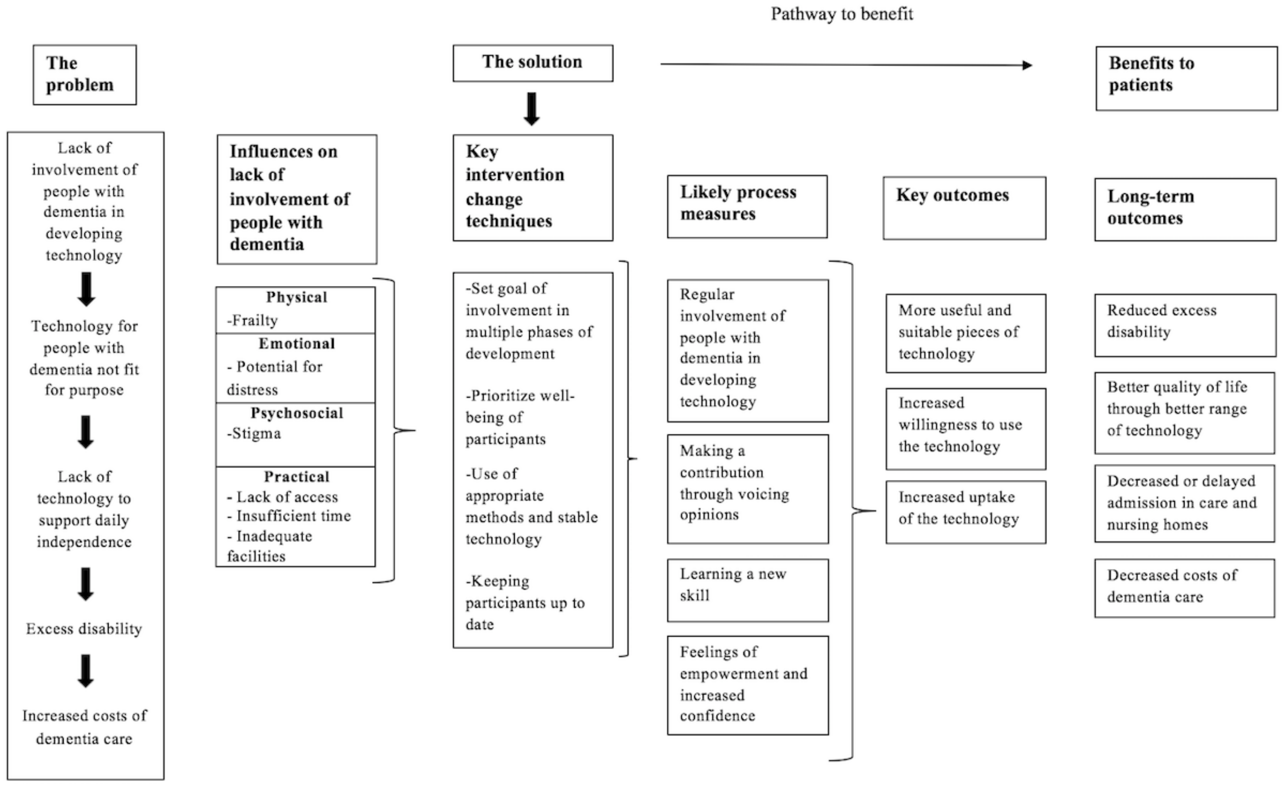
Having the right prerequisites in place before involvement can help overcome the challenges and optimize the involvement of people with dementia. When it comes to the participants, prioritizing their well-being and ensuring that they are aware of the purpose and relevance of their involvement can help contribute to an enjoyable research experience [6,21]. Both the findings from this review and the suggestions from the PPI group members emphasized the need for skilled researchers and the need for a comfortable research environment. Researchers need to take time to get to know participants, and PPI group members added that researchers should be aware of any specialized knowledge of people with dementia before their involvement. This could strengthen their contributions, and it would easily enable them to become coresearchers. Furthermore, determining the goal of involvement and where it is best suited in the development process will help avoid wasting time of people with dementia [27].

Keeping in line with this, multiple methods for involvement need to be considered to obtain the most optimal feedback, and where possible, multiple phases of development should be included. This was confirmed by the PPI group members, and in addition to this, early involvement of people with dementia was considered helpful, as it would also help to identify their own needs and ideas for technology. The latter is crucial in some of the studies included in this review, in which people with dementia are involved in needs assessments and prioritizing areas for functional improvement before moving on to prototype development. It is also recommended to involve all relevant stakeholders and allow interaction between them to obtain a well-rounded view from several user perspectives but also to enable people with dementia to become part of the research and development team [26,28].

During the involvement of people with dementia, the research experience can be further enhanced if participants are able to learn a new skill involving technology [6,18]. This can lead to increased motivation and feelings of empowerment. In addition, the use of appropriate terminology can be helpful in obtaining meaningful and more in-depth answers [30]. Technology must meet an acceptable standard of stability and reliability when evaluating its impact [12]. This can help to avoid frustration among participants and to avoid missing out on essential feedback. PPI group members agreed that it would be more useful to use functional devices during testing and added that the technology should be compatible with different platforms if applicable (eg, a computer or a mobile phone). However, members also reflected on the *Wizard of Oz* method and the idea of an unseen researcher operating the device from a distance while people with dementia would interact with it. This method could potentially function as a good alternative where paper-based prototypes are not suitable and fully functional prototypes are not available. After involvement has taken place, it is advisable to keep participants up to date regarding further development or implementation of the new technology.

Figure 3 includes a logic model based on the findings from this review and the best practice guidelines. It describes the current problem of a lack of involvement of people with dementia in developing technology and how this can be remedied through key intervention change techniques, such as setting goals of involvement and using appropriate methods. This will lead to important short- and long-term outcomes, including more useful pieces of technology and decreased costs of dementia care.

Figure 3. Optimizing the involvement of people with dementia in developing technology-based interventions: logic model.



Limitations (Narrative Synthesis Element 4: Assessing the Robustness of the Synthesis)

This review included very few studies that involved people with dementia in multiple stages of technology development. Furthermore, although this review did not focus on the passive involvement of people with dementia (eg, in large-scale RCTs), few studies allowed for impact evaluation and subsequent sharing of feedback, such as in the study by Hattink et al [12]. Finally, no studies were found that included the involvement of people with dementia in the implementation phase of development.

The definition of involvement in a development process was partly based on previous research and therefore only included studies in which people with dementia played an active part in development or were able to give feedback. This might have caused the exclusion of other potentially relevant studies, which involved people with dementia through other methods, which is a limitation of this review. Another limitation is the focus on English language peer-reviewed journal papers only, which may have led to the exclusion of other potentially relevant content.

Future Research

To develop more tailored technology and explore the possible roles for people with dementia in other phases, future studies should expand on the level of involvement of people with dementia. People with dementia should be coresearchers or advisors and be made an integral part of the research team and the study. This would enable the same group of people with dementia to consistently provide feedback from the early stages of development (eg, formulating the problem) toward the mid stages and end stages (eg, design and implementation). Considering the lack of studies focusing on the implementation phase, future research should explore the role of people with dementia in both the implementation and dissemination of a new technology. In addition, in the majority of the studies, the researcher often acts as a mediator between the person with dementia and the technology developer. However, future studies could aim to facilitate direct knowledge transfer between the two for the technology developers to receive raw feedback.

Conclusions

Over time, studies have involved people with dementia more rigorously in developing technology; however, technologies still need to be more tailored to the needs and preferences of people with dementia. To do this, people with dementia need to be given an active role in the development of technology, so they can have the opportunity to voice their thoughts and opinions. This narrative synthesis review has shown that it is feasible for people with dementia to assume a more active role throughout the development process from discussing and commenting to tryouts and testing. The involvement of people with dementia is associated with several benefits, namely, the development of better and more useful technology, an improved uptake of the technology, and an increased willingness to use the technology. In addition, the evidence-based best practice guidelines were deemed to be relevant by PPI group members and will help support future researchers, technology developers, and people with dementia to optimize involvement when developing technology (Textbox 2). This will not only ensure that future technology-based interventions are suitable but will also allow people with dementia to feel empowered by making an effective contribution to technology development and research in general.

Textbox 2. Summary of best practice guidance for involving people with dementia in developing technology.

Prepare for involvement:

- Make this a positive experience for participants by creating a friendly environment, where people can ask questions and feel supported
- Involve a variety of stakeholders and users to collect a range of feedback and perspectives
- Ensure all practicalities for involvement are in place to meet the needs of participants
- Participants should be made aware of the purpose and relevance of their involvement to meet their expectations and encourage honest feedback
- Explore the available methods for collecting feedback and select the ones best suited for the goal of involvement

Practice involvement:

- Use appropriate terminology and words when asking questions to promote understanding and generate more in-depth feedback
- Offer participants the opportunity to learn a new skill through their involvement to enhance well-being and empowerment
- Involve participants throughout the development process to create a more suitable piece of technology for wider uptake
- Keep participants informed after their involvement so they can stay up to date on further development and implementation of the technology

Acknowledgments

The authors would like to thank the PPI group members at the Institute of Mental Health in Nottingham who gave their time to review the findings of this study and provide essential feedback. The authors would also like to thank Emma Young (information specialist, Duncan Macmillan and Rampton Staff Libraries and Institute of Mental Health) for her support in defining the search strategy and in running the first pilot searches for this review. The research presented in this paper was carried out as part of the Marie Curie Innovative Training Network action, H2020-MSCA-ITN-2015, under grant agreement number 676265.

Authors' Contributions

HR and MO developed the research question for this review. MO provided supervision and guidance throughout the review process. HR conducted the searches, screening for eligibility, data extraction, PPI consultations, and drafted the paper. AB, LY, and JS supported screening for eligibility and data extraction. All authors have read and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Main characteristics of the included studies.
[DOCX File , 23 KB - jmir_v22i12e17531_app1.docx]

Multimedia Appendix 2
Methodological quality of the included qualitative studies.
[DOCX File , 18 KB - jmir_v22i12e17531_app2.docx]

Multimedia Appendix 3
Best practice guidelines for the involvement of people with dementia in developing technology-based interventions.
[DOCX File , 22 KB - jmir_v22i12e17531_app3.docx]

References

1. Prince M, Wimo A, Guerchet M, Ali G, Wu Y, Prina M. World Alzheimer Report 2015. The Global Impact of Dementia: an Analysis of Prevalence, Incidence, Cost and Trends. Alzheimer's Disease International. 2015. URL: https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf [accessed 2018-11-10]
2. Gibson G, Newton L, Pritchard G, Finch T, Brittain K, Robinson L. The provision of assistive technology products and services for people with dementia in the United Kingdom. Dementia (London) 2016 Jul;15(4):681-701. [doi: 10.1177/1471301214532643] [Medline: 24803646]
3. Tak SH, Benefield LE, Mahoney DF. Technology for long-term care. Res Gerontol Nurs 2010 Jan;3(1):61-72 [FREE Full text] [doi: 10.3928/19404921-20091103-01] [Medline: 20128544]

4. Span M, Hettinga M, Vernooij-Dassen M, Eefsting J, Smits C. Involving people with dementia in the development of supportive IT applications: a systematic review. Ageing Res Rev 2013 Mar;12(2):535-551. [doi: 10.1016/j.arr.2013.01.002] [Medline: 23318684]
5. Orpwood R, Chadd J, Howcroft D, Sixsmith A, Torrington J, Gibson G, et al. Designing technology to support quality of life of people with dementia. Technol Disabil 2007;19:103-112 [FREE Full text] [doi: 10.1007/s10209-009-0172-1]
6. Hanson E, Magnusson L, Arvidsson H, Claesson A, Keady J, Nolan M. Working together with persons with early stage dementia and their family members to design a user-friendly technology-based support service. Dementia 2007 Aug 1;6(3):411-434. [doi: 10.1177/1471301207081572]
7. Robinson L, Brittain K, Lindsay S, Jackson D, Olivier P. Keeping in touch everyday (KITE) project: developing assistive technologies with people with dementia and their carers to promote independence. Int Psychogeriatr 2009 Jun;21(3):494-502. [doi: 10.1017/S1041610209008448] [Medline: 19193255]
8. Astell A, Alm N, Gowans G, Ellis M, Dye R, Vaughan P. Involving older people with dementia and their carers in designing computer based support systems: some methodological considerations. Univ Access Inf Soc 2008 May 29;8(1):49-58. [doi: 10.1007/s10209-008-0129-9]
9. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new medical research council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
10. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 5;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
11. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews. Lancaster, UK: Lancaster University; Jan 2006.
12. Hattink BJ, Meiland FJ, Overmars-Marx T, de Boer M, Ebben PW, van Blanken M, et al. The electronic, personalizable Rosetta system for dementia care: exploring the user-friendliness, usefulness and impact. Disabil Rehabil Assist Technol 2016;11(1):61-71. [doi: 10.3109/17483107.2014.932022] [Medline: 24989993]
13. Khosla R, Nguyen K, Chu M. Human robot engagement and acceptability in residential aged care. Int J Hum–Comput Interact 2016 Dec 27;33(6):510-522. [doi: 10.1080/10447318.2016.1275435]
14. Moyle W, Jones C, Dwan T, Petrovich T. Effectiveness of a virtual reality forest on people with dementia: a mixed methods pilot study. Gerontologist 2018 May 8;58(3):478-487. [doi: 10.1093/geront/gnw270] [Medline: 28329867]
15. Moyle W, Jones C, Dwan T, Ownsworth T, Sung B. Using telepresence for social connection: views of older people with dementia, families, and health professionals from a mixed methods pilot study. Aging Ment Health 2019 Dec;23(12):1643-1650. [doi: 10.1080/13607863.2018.1509297] [Medline: 30450924]
16. Begum M, Wang R, Huq R, Mihailidis A. Performance of daily activities by older adults with dementia: the role of an assistive robot. IEEE Int Conf Rehabil Robot 2013 Jun;2013:6650405. [doi: 10.1109/ICORR.2013.6650405] [Medline: 24187224]
17. Freeman ED, Clare L, Savitch N, Royan L, Litherland R, Lindsay M. Improving website accessibility for people with early-stage dementia: a preliminary investigation. Aging Ment Health 2005 Sep;9(5):442-448. [doi: 10.1080/13607860500142838] [Medline: 16024403]
18. Span M, Hettinga M, Groen-van de Ven L, Jukema J, Janssen R, Vernooij-Dassen M, et al. Involving people with dementia in developing an interactive web tool for shared decision-making: experiences with a participatory design approach. Disabil Rehabil 2018 Jun;40(12):1410-1420. [doi: 10.1080/09638288.2017.1298162] [Medline: 28286969]
19. Orpwood R, Chadd J, Howcroft D, Sixsmith A, Torrington J, Gibson G, et al. Designing technology to improve quality of life for people with dementia: user-led approaches. Univ Access Inf Soc 2009 Oct 27;9(3):249-259. [doi: 10.1007/s10209-009-0172-1]
20. Klein P, Uhlig M, Will H. The touch and feel of the past—using Haptic and VR artefacts to enrich reminiscence therapy for people with dementia. Technologies 2018 Nov 13;6(4):104. [doi: 10.3390/technologies6040104]
21. Boman I, Nygård L, Rosenberg L. Users' and professionals' contributions in the process of designing an easy-to-use videophone for people with dementia. Disabil Rehabil Assist Technol 2014 Mar;9(2):164-172. [doi: 10.3109/17483107.2013.769124] [Medline: 24512219]
22. Lopes P, Pino M, Carletti G, Hamidi S, Legué S, Kerhervé H, et al. Co-conception process of an innovative assistive device to track and find misplaced everyday objects for older adults with cognitive impairment: the Trouve project. IRBM 2016 Apr;37(2):52-57. [doi: 10.1016/j.irbm.2016.02.004]
23. McCabe L, Innes A. Supporting safe walking for people with dementia: user participation in the development of new technology. Gerontechnology 2013;12(1):4-15 [FREE Full text] [doi: 10.4017/gt.2013.12.1.006.00]
24. Meiland FJ, Hattink BJ, Overmars-Marx T, de Boer ME, Jedlitschka A, Ebben PW, et al. Participation of end users in the design of assistive technology for people with mild to severe cognitive problems; the European Rosetta project. Int Psychogeriatr 2014 May;26(5):769-779. [doi: 10.1017/S1041610214000088] [Medline: 24507571]

25. Davies R, Nugent C, Donnelly M, Hettinga M, Meiland F, Moelaert F, et al. A user driven approach to develop a cognitive prosthetic to address the unmet needs of people with mild dementia. Pervasive Mob Comput 2009 Jun;5(3):253-267. [doi: 10.1016/j.pmcj.2008.07.002]
26. Kerkhof YJ, Rabiee F, Willems CG. Experiences of using a memory aid to structure and support daily activities in a small-scale group accommodation for people with dementia. Dementia (London) 2015 Sep;14(5):633-649. [doi: 10.1177/1471301213504210] [Medline: 24339118]
27. Martin S, Augusto JC, McCullagh P, Carswell W, Zheng H, Wang H, et al. Participatory research to design a novel telehealth system to support the night-time needs of people with dementia: nocturnal. Int J Environ Res Public Health 2013 Dec 04;10(12):6764-6782 [FREE Full text] [doi: 10.3390/ijerph10126764] [Medline: 24304507]
28. Jamin G, Luyten T, Delsing R, Braun S. The process of co-creating the interface for VENSTER, an interactive artwork for nursing home residents with dementia. Disabil Rehabil Assist Technol 2018 Nov;13(8):809-818. [doi: 10.1080/17483107.2017.1385102] [Medline: 29037109]
29. Meiland FJ, Bouman AI, Sävenstedt S, Bentvelzen S, Davies RJ, Mulvenna MD, et al. Usability of a new electronic assistive device for community-dwelling persons with mild dementia. Aging Ment Health 2012;16(5):584-591. [doi: 10.1080/13607863.2011.651433] [Medline: 22360649]
30. Topo P, Mäki O, Saarikalle K, Clarke N, Begley E, Cahill S, et al. Assessment of a music-based multimedia program for people with dementia. Dementia 2004;3(3):331-350. [doi: 10.1177/1471301204045164]
31. CASP Checklists. CASP: Critical Appraisal Skills Programme. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2019-05-29]
32. Bayliss A, Currie L, McIntosh T, Bazian L, Gallagher R, Watson M. Infusion Therapy Standards: Rapid Evidence Review. Royal College of Nursing. 2016. URL: https://www.rcn.org.uk/-/media/royal-college-of-nursing/documents/publications/2016/december/005702.pdf?la=en [accessed 2018-09-14]
33. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
34. McDermott O, Crellin N, Ridder HM, Orrell M. Music therapy in dementia: a narrative synthesis systematic review. Int J Geriatr Psychiatry 2013 Aug;28(8):781-794. [doi: 10.1002/gps.3895] [Medline: 23080214]
35. di Lorito C, Birt L, Poland F, Csipke E, Gove D, Diaz-Ponce A, et al. A synthesis of the evidence on peer research with potentially vulnerable adults: how this relates to dementia. Int J Geriatr Psychiatry 2017 Jan;32(1):58-67. [doi: 10.1002/gps.4577] [Medline: 27633160]

Abbreviations

CASP: Critical Appraisal Skills Programme
CeHRes: Centre for eHealth Research
CINAHL: Cumulated Index to Nursing and Allied Health Literature
EMBASE: Excerpta Medica database
MEDLINE: Medical Literature Analysis and Retrieval System Online
MRC: Medical Research Council
PPI: patient and public involvement
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
VRF: virtual reality forest

Edited by G Eysenbach; submitted 18.12.19; peer-reviewed by M Mulvenna, G Gibson; comments to author 15.02.20; accepted 26.02.20; published 03.12.20.

Please cite as:
Rai HK, Cavalcanti Barroso A, Yates L, Schneider J, Orrell M
Involvement of People With Dementia in the Development of Technology-Based Interventions: Narrative Synthesis Review and Best Practice Guidelines
J Med Internet Res 2020;22(12):e17531
URL: https://www.jmir.org/2020/12/e17531
doi:10.2196/17531
PMID:33270034

Review

Visual Analytic Tools and Techniques in Population Health and Health Services Research: Scoping Review

Jawad Ahmed Chishtie[1,2,3,4], MSc, MD; Jean-Sebastien Marchand[5], PhD; Luke A Turcotte[2,6], PhD; Iwona Anna Bielska[7,8], PhD; Jessica Babineau[9], MLIS; Monica Cepoiu-Martin[10], PhD, MD; Michael Irvine[11,12], PhD; Sarah Munce[1,4,13,14], PhD; Sally Abudiab[1], BSc; Marko Bjelica[1,4], MSc; Saima Hossain[15], BSc; Muhammad Imran[16], MSc; Tara Jeji[3], MD; Susan Jaglal[15], PhD

[1]Rehabilitation Sciences Institute, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[2]Advanced Analytics, Canadian Institute for Health Information, Toronto, ON, Canada
[3]Ontario Neurotrauma Foundation, Toronto, ON, Canada
[4]Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada
[5]Universite de Sherbrooke, Quebec, QC, Canada
[6]School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada
[7]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada
[8]Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, ON, Canada
[9]Library & Information Services, University Health Network, Toronto, ON, Canada
[10]Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[11]Department of Mathematics, University of British Columbia, Vancouver, BC, Canada
[12]British Columbia Centre for Disease Control, Vancouver, BC, Canada
[13]Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, ON, Canada
[14]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[15]Department of Physical Therapy, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[16]Department of Epidemiology and Public Health, Health Services Academy, Islamabad, Pakistan

Corresponding Author:
Jawad Ahmed Chishtie, MSc, MD
Rehabilitation Sciences Institute
Faculty of Medicine
University of Toronto
851, 500 University Ave
Toronto, ON, M5G 1V7
Canada
Phone: 1 416 946 7902
Email: jac161@gmail.com

Abstract

Background: Visual analytics (VA) promotes the understanding of data with visual, interactive techniques, using analytic and visual engines. The analytic engine includes automated techniques, whereas common visual outputs include flow maps and spatiotemporal hot spots.

Objective: This scoping review aims to address a gap in the literature, with the specific objective to synthesize literature on the use of VA tools, techniques, and frameworks in interrelated health care areas of population health and health services research (HSR).

Methods: Using the 2018 PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, the review focuses on peer-reviewed journal articles and full conference papers from 2005 to March 2019. Two researchers were involved at each step, and another researcher arbitrated disagreements. A comprehensive abstraction platform captured data from diverse bodies of the literature, primarily from the computer and health sciences.

Results: After screening 11,310 articles, findings from 55 articles were synthesized under the major headings of visual and analytic engines, visual presentation characteristics, tools used and their capabilities, application to health care areas, data types and sources, VA frameworks, frameworks used for VA applications, availability and innovation, and co-design initiatives. We found extensive application of VA methods used in areas of epidemiology, surveillance and modeling, health services access, use, and cost analyses. All articles included a distinct analytic and visualization engine, with varying levels of detail provided. Most tools were prototypes, with 5 in use at the time of publication. Seven articles presented methodological frameworks. Toward consistent reporting, we present a checklist, with an expanded definition for VA applications in health care, to assist researchers in sharing research for greater replicability. We summarized the results in a Tableau dashboard.

Conclusions: With the increasing availability and generation of big health care data, VA is a fast-growing method applied to complex health care data. What makes VA innovative is its capability to process multiple, varied data sources to demonstrate trends and patterns for exploratory analysis, leading to knowledge generation and decision support. This is the first review to bridge a critical gap in the literature on VA methods applied to the areas of population health and HSR, which further indicates possible avenues for the adoption of these methods in the future. This review is especially important in the wake of COVID-19 surveillance and response initiatives, where many VA products have taken center stage.

International Registered Report Identifier (IRRID): RR2-10.2196/14019

(J Med Internet Res 2020;22(12):e17892) doi:10.2196/17892

KEYWORDS

visual analytics; machine learning; data visualization; data mining; population health; health services research; mobile phone

Introduction

Background

Visual analytics (VA) is a term that was formally introduced in the literature 15 years ago [1,2]. It describes a semiautomated approach to electronic data processing, guided by users who are able to interact with data through an interface [3,4]. In essence, VA transforms large amounts of quantitative or qualitative information into graphical formats that can be modified based on the operator's needs [4,5]. The resulting views can be used by users with diverse backgrounds to better understand data, communicate results, and disseminate information across a broad spectrum of disciplines [6,7].

The implementation and use of VA have bloomed in many sectors of health care systems during the past decade [8]. Population health research involves the study of data related to health outcomes and determinants among and between populations [9,10], whereas health services research (HSR) explores the functioning of the health care system and its workforce in relation to access, quality, costs, and patient outcomes [11,12]. Both fields involve the analysis of big data, including information collected through clinical databases, administrative data sets, or electronic health records (EHRs) [13-15]. VA offers the opportunity for health data users, such as clinicians, researchers, decision makers, and consumers, to visually explore and interpret complex data sets to guide decision making and knowledge discovery [3,16].

Rationale

Although researchers have pointed out the lack of literature on the extent of the use of VA applications in various sectors [3], we identified 4 recent systematic reviews that covered varied areas of VA applications in health care. The 2018 paper by Islam et al [17] was one of the most comprehensive reviews about data mining applications in health care. However, the review is limited to mining approaches for health care data and does not primarily cover VA. The recently published review by Chung et al [8] relates to VA approaches in mental health care systems and policy. One of the most cited systematic reviews is that by West et al [18] on the use of visualization for EHRs aimed at knowledge discovery. Although these reviews cover some aspects of the wide field of visualization and analytics in health care, none have focused on areas of population health and HSR.

One of the seemingly close literature syntheses is the review by Wu et al [19] on visualization and VA technologies in medical informatics for characterizing evaluation methods. However, there are significant distinctions between that paper and our review. First, their review [19] relates to the subject area of health informatics, which is almost exclusively concerned with patient data in the context of care provision. The classic definition of the subject area is "the applications of information technology to healthcare delivery" [20]. Second, Wu et al [19] cover evaluation methods for VA applications and not VA applications themselves. Our scoping review focuses on methods related to VA applications in population health and HSR and does not focus on evaluation methods.

Through this review, we attempt to bridge a critical gap in the literature on the use of VA tools, techniques, and frameworks in the interrelated and overlapping areas of population health and HSR. To the best of our knowledge, none of the recent systematic literature syntheses focused on these areas of health care or covered the VA tools and techniques that we present in this scoping review.

In response to this broader conceptualization, this scoping review identified and synthesized findings from English language peer-reviewed sources that used VA approaches and methods in population health and HSR. Such a synthesis of the literature will be helpful for researchers, practitioners, and decision support analysts to (1) explore recent trends in the use of innovative VA methods in the important health care domains of population health and HSR, (2) learn from methodological frameworks, and (3) uptake these techniques to meet the growing needs for data-driven insights. Furthermore, this review presents the settings for which VA applications are developed and applied as well as the intended target audience. This information is important in the context of the use of VA techniques in participatory co-design initiatives.

Objectives

The objectives of this review are (1) to identify the scope and nature of the use of VA methods in population health and HSR and (2) to summarize methodological tools, techniques, and frameworks from peer-reviewed literature in both health care areas.

Methods

Protocol and Overall Scoping Review Methodology

The study protocol was previously published, detailing the search strategy and methods [21]. We primarily followed the Joanna Briggs Institute guidelines on scoping reviews [22] and the framework by Arksey and O'Malley for conducting scoping reviews [23], with improvements suggested by Levac et al [24] and Peters et al [25] for conceptualizing the population, concepts, and context of the study, especially given the context of a methods-based review.

We further used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) checklist from the work by Tricco et al [26] to operationalize the different steps, while providing milestones and guideposts for adaptation during the review. The checklist is shown in Multimedia Appendix 1. We followed the journal guidelines for the preparation of the manuscript. The major methodological steps for the systematic scoping review comprised determining the research question; identifying relevant studies; title, abstract, and full-text screening; data abstraction; and the collation, summarization, and reporting of the results.

Eligibility Criteria

The inclusion and exclusion criteria are presented in Textboxes 1 and 2, respectively. Papers included during the screening stage needed to have a central VA component with a focus on population health or HSR. Studies conducted in clinical settings or focusing on a single condition, without a population or health service component, were not included in the review. The operational definitions for all concepts are presented in detail later in this section.

Textbox 1. Inclusion criteria for selection of articles.

Inclusion criteria

- Peer-reviewed or conference papers
- January 1, 2005, to March 31, 2019
- Population health or health services research (HSR) related
- Articles with population-level or HSR metrics: incidence, prevalence, events over time and space, spatiotemporal, access, utilization, disease or condition distribution, and social or multiple determinants of health
- Articles with an analytic engine and a visualization engine
- Articles with exploratory data analytic techniques
- Articles on electronic medical records and electronic health records
- Articles with dashboards with an explicit analytic engine to feed data
- Articles with automated analysis, data mining techniques, interactive tools, and iterative analysis

Textbox 2. Exclusion criteria for the articles.

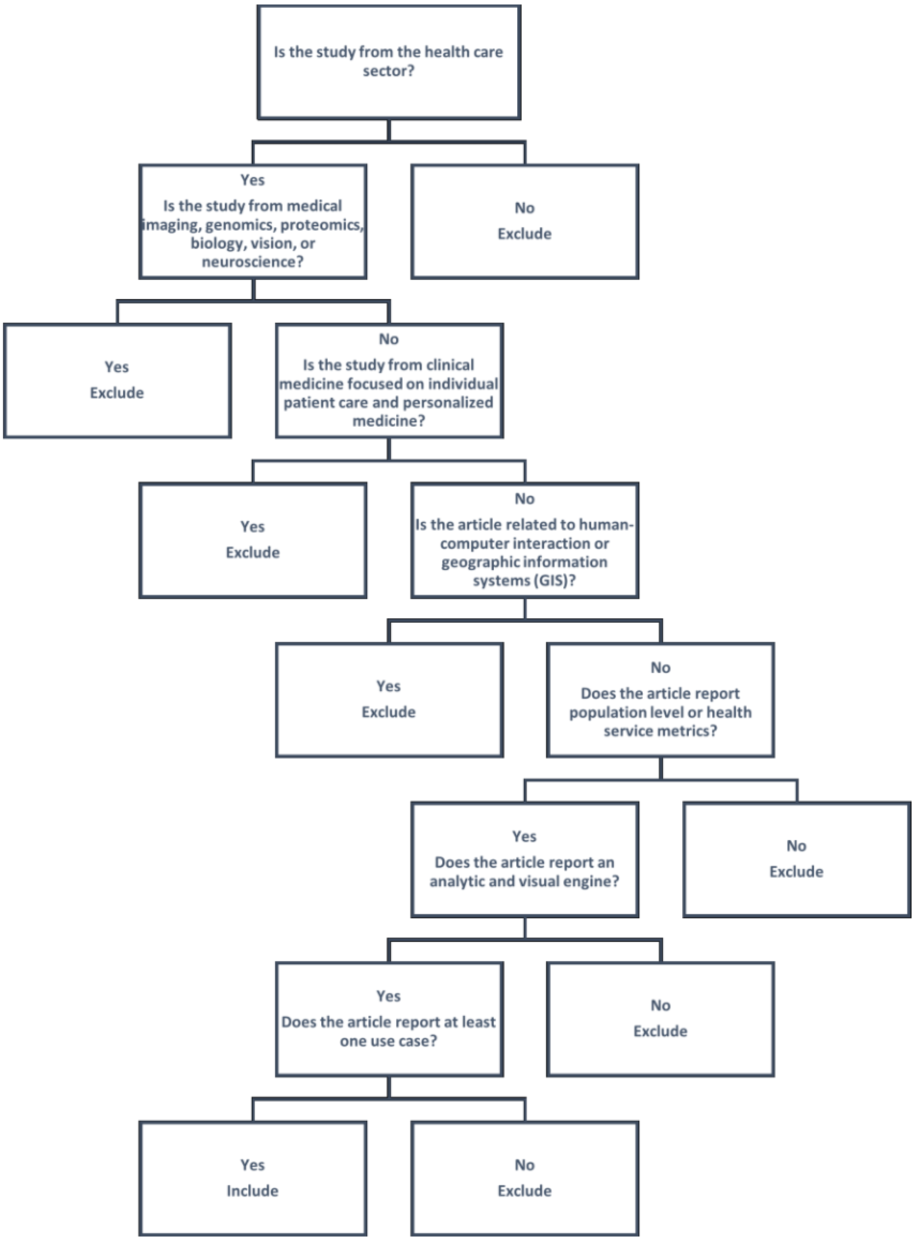
Exclusion criteria

- Articles not in the English language
- Editorials, projects, reviews, book chapters, short papers, or reports
- Articles on medical imaging
- Studies conducted in clinical settings without a population-level or health services research component
- Articles for individual condition from a single hospital or unit, such as intensive care, surgery, anesthesia, without a population-level or health services research (HSR) component
- Articles on device or sensor data without a population-level or HSR component
- Studies lacking an analytic method or engine
- Cartographic or geographic information systems (GIS) method

One of the primary aims of both population health and HSR is to better understand disease distribution and barriers to equitable care. We included these components and related metrics for population health, such as incidence, prevalence, and events over time and space, to guide us in delineating research that focused on clinical or individual conditions or cases. For example, if a diabetes dashboard presented clinical care for an individual in a hospital setting, such as blood sugar or glycated hemoglobin levels, it was excluded. However, if a diabetes dashboard presented a *population* with glycated hemoglobin levels in a hospital catchment area, it was included as it had a population-level component. Studies without an analytic engine were excluded. Finally, articles not in the English language and non–peer-reviewed work, such as editorials, projects, short papers, conference abstracts, and reports, were excluded.

The eligibility criteria were revised twice during the screening process. In total, 4 items were added later to the exclusion criteria: studies conducted in clinical settings without a population-level component, articles on device or sensor data, articles related to cartographic methods, and articles related to geographic information systems (GIS) techniques. However, VA articles with a GIS component covering spatiotemporal data, sometimes termed geo-VA, were included in the review. Figure 1 shows a simplified decision tree for the selection of articles.

Figure 1. Decision tree for assessing article eligibility.



Scoping Review Timeline

We were able to trace the first formal use of the term VA to the seminal work by Thomas and Cook in 2005 [2,3]. As the use of the term was in the area of national defense, we expected a lag time in the adoption of the methodology and the use of the term in health care. On the basis of these reasons, our multidisciplinary team decided to include articles from January 1, 2005, to March 31, 2019.

Information Sources

The full electronic search strategy is provided in Multimedia Appendix 2. It was developed through an iterative process by the research team, which included an information specialist (JB). A preliminary search was conducted in MEDLINE, following which the first 100 resulting article abstracts were reviewed to refine the search strategy. The search strategy was then peer reviewed using the Peer Review of Electronic Search Strategies (PRESS) guidelines by a second information specialist. A total of 6 databases were searched in April 2019 using both keywords and subject-specific vocabulary (eg, Medical Subject Headings [MeSH], Emtree; Table 1). We have detailed the search strategy, the keywords used, and the operationalization of the concepts in depth in the published protocol [21].

Table 1. Databases and search results.

Database	Platform	Search results (n=14,099), n
MEDLINE	OvidSP	4633
EMBASE	OvidSP	1880
Web of Science core collection	Web of Science	5396
Compendex	Engineering Village	1267
IEEE Xplore	IEEE	151
Inspec	Engineering Village	772

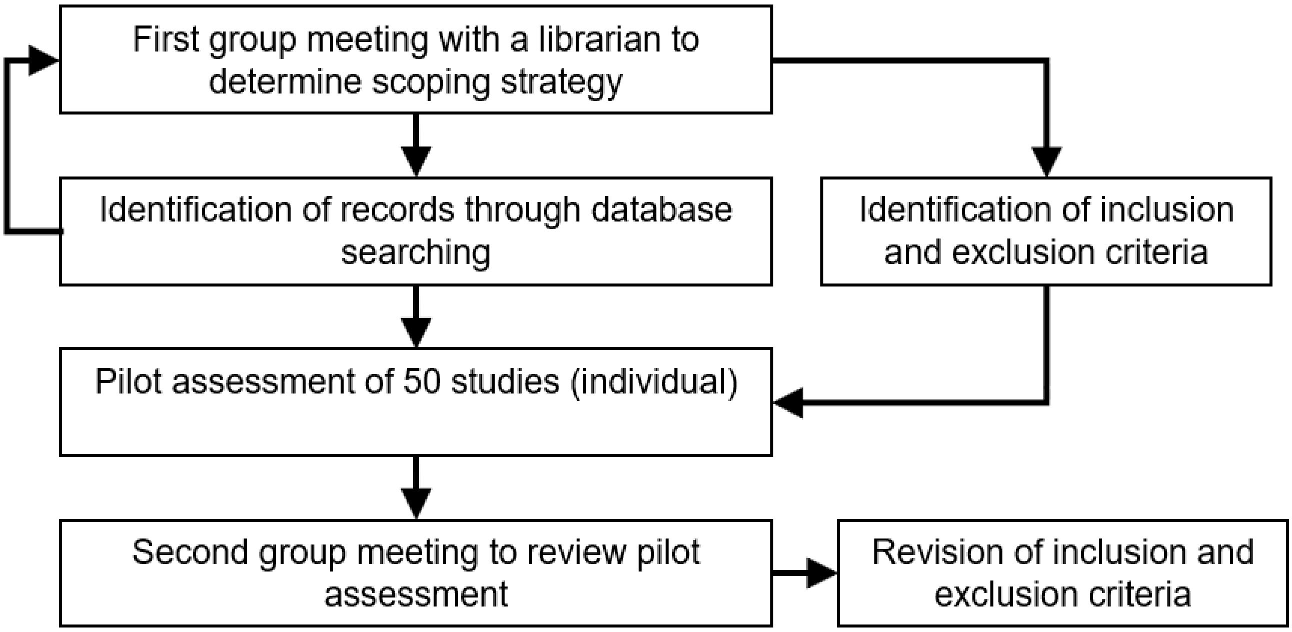
The review management software Covidence was used to manage the search results, including the importing of references, screening of citations, and conflict resolution [27]. Duplicates were removed in 3 phases. First, citations were checked in EndNote (Clarivate Analytic) for duplicates, followed by duplicate identification by Covidence (Veritas Health Innovation) systematic review software. Finally, duplicates were removed manually during the full-text review.

To complement the database searches, we conducted an internet search using Google and Google Scholar search engines, and we manually searched 10 journals deemed relevant to the research question. These were *Applied Clinical Informatics*, *Visual Analytics in Healthcare Proceedings*,*IEEE Transactions on Information Technology in Biomedicine*, *Journal of Medical Internet Research*, *Journal of Medical Systems*, *Journal of the American Medical Informatics Association*, *Health Affairs*, *Journal of Biomedical Informatics*, *Healthcare Informatics Research*, and *PLOS One*. We further reviewed the conference proceedings from Visual Analytics in Health Care, which is held one year apart in collaboration with the American Medical Information Association and IEEE VIS conferences. In addition, we reviewed the references from another 13 systematic and narrative topic-related reviews identified during the screening of the articles [7,16-19,28-35].

Selection of Sources of Evidence

The process for the selection of sources of evidence was divided into 2 phases. First, to enhance the consistency among reviewers in the team, we met to discuss the inclusion and exclusion criteria. We randomly selected 50 articles that each reviewer screened for title and abstract. After this initial pilot assessment, we discussed the process, criteria, conflicts, ambiguities, and difficulties encountered. This pilot phase led to a slight readjustment of the inclusion and exclusion criteria. This iterative methodology, with the selection of sources of evidence is illustrated in Figure 2. In the second phase, 2 reviewers were required for the title and abstract screening process as well as for the full-text screening process. In both cases, conflicts were resolved by another reviewer.

Figure 2. Pilot assessment and revision of criteria for selection of sources of evidence.



Data Charting

The data charting form was developed specifically for this review and piloted with 3 randomly selected articles to refine the categories of abstraction. Each article was then assigned to 2 independent researchers. A third researcher offered arbitration, correction, and validation of the abstraction where required. For the specific abstraction fields, we followed the definitions detailed in the *Operational Concepts and Definitions* section.

Data Items and Synthesis of Results

Data abstraction was based on 5 major categories: (1) study characteristics (eg, country, problem, settings, target audience); (2) frameworks, tools, and techniques used (eg, tool name, framework followed); (3) analytic and visualization methods and engines used (eg, analytic engine, data type, data used); (4) domains of health care and type of measures used (eg, population health, health services); and (5) study innovation, impact, availability of the tool, and whether it was co-designed with the target audience. The results were compiled into tables under these major categories, following the major schema developed during the operationalization of the concepts and abstraction of the articles.

Operational Concepts and Definitions

Our initial literature search revealed varying definitions and inconsistent use of terms for the 3 major concepts: population health, HSR, and VA. To translate these terms into operational definitions, we undertook a 3-pronged strategy of studying seminal literature, recent systematic reviews, and subject trees in MEDLINE. Although the search terms and their sources are detailed in the study protocol [21], we detail the operationalized concepts below.

Population Health and HSR

Defined as the “science and art of preventing disease” [36], population health is nested under the larger concept of public health. Although experts have attempted to develop a common language related to public and population health [37], population health as a MeSH term was only recently added to MEDLINE in 2018 [38]. Kindig and Stoddart [9,10] define population health as “the health outcomes of a group of individuals, including the distribution of such outcomes within the group” that includes “health outcomes, patterns of health determinants, and policies and interventions that link these two.” We found the expanded definition of population health by Kindig and Stoddart [9,10] apt for our review to encompass the vast nature of this area. To adapt search terms, we studied the national public health language created by the National Institute for Health and Care Excellence in the United Kingdom [39], detailed database trees, and 5 recent reviews [40-44].

HSR is defined by the Canadian Institutes of Health Research as research with the “goal of improving the efficiency and effectiveness of health professionals and the health care system” [11]. Population health and HSR are intertwined concepts, with overlapping research and communities of practice, first, in the purview of studying problems through a population lens and, second, through a health systems and service lens. The *population health approach* brings together the two in their application toward health sector reform, allowing researchers to formulate proposals for the organization and delivery of health care [41,45].

Guided by 4 recent reviews [45-48] and the filters for HSR developed by the National Library of Medicine [49], we translated the concept to the search strategy. Particular to HSR, we included studies on access, utilization, and cost of health services in the review.

VA: Analytic and Interactive Visual Engines

The seminal work by Thomas and Cook [2] defines VA as “the science of analytical reasoning facilitated by interactive visual interfaces.” Later, Keim et al [3] extended this concept to “automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets.” Although these definitions offered a high-level conceptualization of the expansive field of VA, we needed a simplified, more encompassing conceptual definition that could help contextualize VA methods and applications in health care. Hence, we opted to use the expanded definition of VA applications in health care by Ola and Sedig [50], comprising analytic and interactive visualization engines. Typically, the analytics engine involves data storage, transformation, and analysis, whereas the visualization engine provides functionality toward data manipulation and display [50].

The analytics engine can employ advanced statistical and machine learning (ML) techniques for various functions. For example, an extract, transform, and load engine using ML algorithms can bring together a database that the visual engine uses to produce visualizations [50]. For the purposes of the review and its focus on population health and HSR, we avoided the term artificial intelligence.

ML is a subset of artificial intelligence methods that includes fitting models to data and learning by training models with data [51]. We focused on tasks such as clustering, classification, and algorithms used to present the major techniques used toward the analytic engine.

Interactivity is one of the recent hallmarks of VA applications, owing to the manipulation of visual interfaces afforded by computing power [50]. We borrow from works by Ola and Sedig [50] and Pike et al [52] to define *interactivity* as the ability to reflect changes in the visual representation of data based on one or more indicators available on the analytic interface to the user. Pike et al [52] categorize interaction elements into 2 main types: *lower level* aimed at change of the visual representation to study patterns, relationships, and the like and *higher level* that offers an understanding of the intent of interaction itself toward knowledge discovery. To select the appropriate literature as part of this scoping review, we focused on lower-level interaction that includes tasks such as filtering, determining ranges, finding anomalies, clustering, and the like by providing menus, dropdowns, and other options on the visualization interface.

Furthermore, to operationalize the search terms related to VA, we studied 4 recent reviews [7,17-19] in addition to 9 seminal papers [6,16,53-59].

Analytic Types and Capability, Settings, and Target Audience

To operationalize the types of analytics that the application targeted within the use case, we adapted the work by Islam et al [17] on data mining techniques in health care. Analytics is defined as "knowledge discovery by analyzing, interpreting and communicating data" [17].

Related to the analytic capability, applications were categorized as being primarily meant for descriptive, predictive, or prescriptive analytics for visual exploration of complex data sets or a combination of these. Descriptive analytics is defined as "exploration and discovery of information in the dataset," predictive analytics is defined as "prediction of upcoming events based on historical data," and prescriptive analytics is defined as "utilization of scenarios to provide decision support" [17]. Although the visual exploration of complex data sets can be seen as an extension of descriptive analytic capability, we kept it as a separate category.

We gleaned information from different parts of the included articles to obtain the study setting and audience based on the potential application for the method, tool, or its user, as mentioned by the authors.

Tools, Applications, and Frameworks

Tools were defined as software used to develop an application to address a certain problem, whereas applications were one or more software program using code or front-end programming employed for data analysis and visualization.

Frameworks in research form the foundation, backbone, or the *blueprint* on which knowledge is constructed [60]. Hence, we opted to study the frameworks that formed the basis for the applications to better situate the literature on VA and to help define the lens, perspective, and conceptual background for the methods. We defined a framework as an extension of a lens or perspective of inquiry that is structured to allow methodological uniformity, adaptation, reporting, understanding, and replicability. Given that our review is methods based, we did not differentiate between a theoretical or conceptual framework [60].

Use Case and Data Source

A use case is defined as the application of the method to an actual data set, source, or simulation data related to population health or HSR. We studied whether the use case included a single data source or multiple data sources. The goal of the application was ascertained by studying whether the application, tool, or method was meant for decision support, knowledge discovery, or both.

Domains of Health Care

Finally, we adapted the domains of health care from Islam et al [17] to represent population health, HSR, or both. We further divided population health–related articles into clinical, demographic, epidemiologic, spatiotemporal, or a combination of these categories. The clinical category would include a condition, the demographic category would include any population-related characteristic such as age, the epidemiologic category would include disease distribution and dynamics, and the spatiotemporal category would include events over time and space. An overlap between the categories within the articles was expected.

Co-Design and Knowledge Translation

Knowledge translation is a wide term used in different contexts, focusing on the translation of research evidence to policies and practice [61]. Although our initial conceptualization for the review was related to knowledge co-creation for decision making, we realized that for the purpose of this review, a better approach would be to consider *co-design* methods especially in the development stages of an application. We used the definition of co-design in health care by Ward et al [62] that encompasses the partnership of health workers, patients, and designers who aspire to change, depending on shared knowledge to achieve *better outcomes or improved efficiency*.

Co-designed applications would have better viability and uptake toward both knowledge transfer and decision support. We studied whether the authors involved stakeholders or target audiences during the development of an application. We did not study co-design methods and approaches, as this was not the objective of the review.

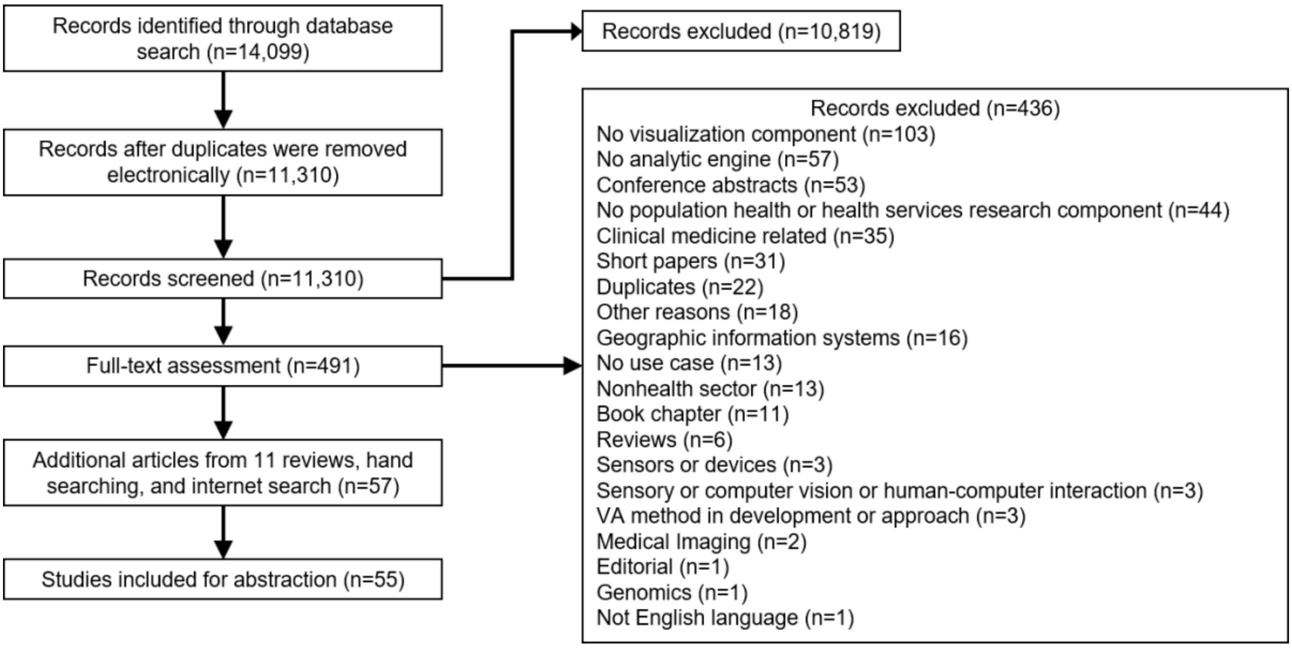
Results

Selection of Articles

We identified 14,099 articles through the combined database searches. Using EndNote, 2078 duplicates were electronically removed in 6 iterations run on 2 different versions, X7 and X9. On importing 12,021 records into Covidence, another 711 duplicates were removed. We screened the titles and abstracts for 11,310 records, of which 10,819 (95.65%) were excluded. We were able to identify 57 more references from 4 systematic reviews identified during the screening process [8,17-19] and hand searching. The results are summarized in a Tableau dashboard [63].

Of the 491 records included for full-text assessment, 436 (88.8%) were excluded. Reasons for exclusion were lacking a visualization component (n=103), lacking an analytic engine (n=57), conference abstracts and editorials (n=53), not population health or HSR (n=44), clinical medicine related (n=35), short papers (n=31), duplicates (n=22), cartographic and GIS methods (n=16), no use case (n=13), non–health sector (n=13), book chapter (n=11), reviews (n=6), sensors or devices (n=3), sensory or computer vision or human-computer interaction (n=3), VA method in development or approach (n=3), medical imaging (n=2), genomics (n=1), not English language (n=1), and other reasons (n=18). Overall, 55 articles were included for abstraction. The PRISMA-ScR flow diagram is shown in Figure 3.

Figure 3. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) chart for article selection. VA: visual analytics.



Study Characteristics, Settings, and Target Audience

The country of the study was ascertained from the actual use case of the VA application. The 55 studies included in the analysis were from 19 countries, including the United States (24/55, 44%), Canada (5/55, 9%), and Germany (3/55, 5%). Details are provided in Multimedia Appendix 3 [64-118].

In terms of the settings where the research took place, studies were most often conducted as part of, or with the involvement of, a government unit, including a ministry or health department (38/55, 69%), followed by academic settings (35/55, 64%), mixed government and academic settings (18/55, 33%), and the industry (3/55, 5%). The intended target audience was the population health and HSR community (53/55, 96%), academic researchers and data scientists (47/55, 85%), clinicians (21/55, 38%), both clinicians and population health and HSR practitioners (21/55, 38%), policy and decision makers (7/55, 13%), consumers and the general public (5/55, 9%), and the industry (3/55, 5%). Table 2 details the study setting, while Table 3 details the target audience in the included papers.

Table 2. Setting of the use cases.

Setting	Values, n (%)	Study (reference)
Academic	38 (69)	Abusharekh et al, 2015 [67]; Afzal et al, 2011 [85]; Ali et al, 2016 [68]; Alonso et al, 2012 [92]; Antoniou et al, 2010 [93]; Benis et al, 2017 [89]; Bryan et al, 2015 [64]; Byrd et al, 2016 [94]; Chorianopoulos et al, 2016 [96]; Garcia-Marti et al, 2017 [97]; Gotz et al, 2014 [76]; Guo et al, 2007 [69]; Hardisty et al, 2010 [100]; Hund et al, 2016 [90]; Ji et al, 2012 [102]; Ji et al, 2013 [81]; Jiang et al, 2016 [103]; Kaieski et al, 2016 [104]; Katsis et al, 2017 [105]; Kostkova et al, 2014 [75]; Lavrac et al, 2007 [70]; Lu et al, 2017 [71]; Luo et al, 2016 [78]; Maciejewski et al, 2010 [107]; Maciejewski et al, 2011 [79]; Marek et al, 2015 [108]; Ozkaynak et al, 2015 [111]; Park et al, 2018 [112]; Perer et al, 2015 [113]; Proulx et al, 2006 [114]; Shaban-Nejad et al, 2017 [84]; Tate et al, 2014 [87]; Widanagamaachchi et al, 2017 [72]; Xing et al, 2010 [91]; Xu et al, 2013 [73]; Yan et al, 2013 [118]; Yu et al, 2017 [82]; Yu et al, 2018 [74]
Government, ministry, or health department	35 (64)	Abusharekh et al, 2015 [67]; Afzal et al, 2011 [85]; Alonso et al, 2012 [92]; Antunes de Mendonca et al, 2015 [86]; Baytas et al, 2016 [80]; Benis et al, 2017 [89]; Bryan et al, 2015 [64]; Castronovo et al, 2009 [77]; Chen et al, 2016 [95]; Dagliati et al, 2018 [66]; Deodhar et al, 2015 [65]; Gligorijevi et al, 2017 [98]; Haque et al, 2014 [99]; Hardisty et al, 2010 [100]; Huang et al, 2015 [101]; Jiang et al, 2016 [103]; Jinpon et al, 2017 [83]; Kaieski et al, 2016 [104]; Kruzikas et al, 2014 [106]; Lavrac et al, 2007 [70]; Lu et al, 2017 [71]; Maciejewski et al, 2011 [79]; Mitrpanont et al, 2017 [109]; Mittelstadt et al, 2014 [110]; Ozkaynak et al, 2015 [111]; Proulx et al, 2006 [114]; Shaban-Nejad et al, 2017 [84]; Soulakis et al, 2015 [115]; Tilahun et al, 2014 [88]; Toddenroth et al, 2014 [116]; Torres et al, 2012 [117]; Xu et al, 2013 [73]; Yan et al, 2013 [118]; Yu et al, 2017 [82]; Yu et al, 2018 [74]
Academic and government or ministry or health department	18 (33)	Abusharekh et al, 2015 [67]; Afzal et al, 2011 [85]; Alonso et al, 2012 [92]; Benis et al, 2017 [89]; Bryan et al, 2015 [64]; Hardisty et al, 2010 [100]; Jiang et al, 2016 [103]; Kaieski et al, 2016 [104]; Lavrac et al, 2007 [70]; Lu et al, 2017 [71]; Maciejewski et al, 2011 [79]; Ozkaynak et al, 2015 [111]; Proulx et al, 2006 [114]; Shaban-Nejad et al, 2017 [84]; Xu et al, 2013 [73]; Yan et al, 2013 [118]; Yu et al, 2017 [82]; Yu et al, 2018 [74]
Industry	3 (5)	Gotz et al, 2014 [76]; Perer et al, 2015 [113]; Yu et al, 2018 [74]

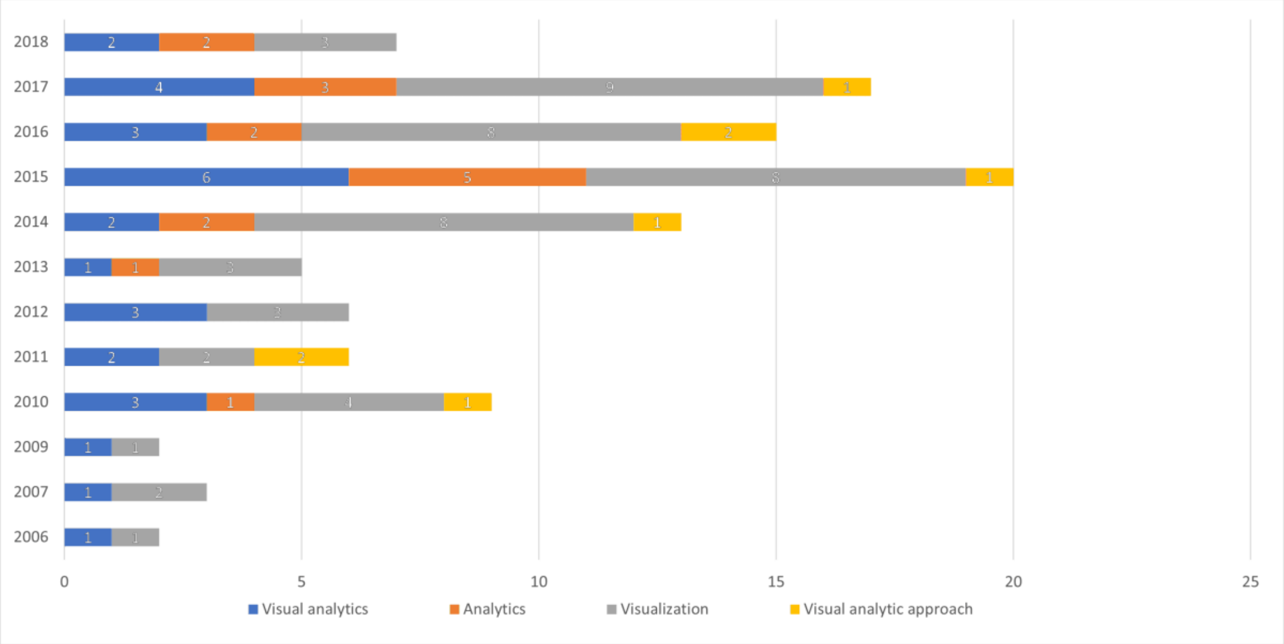
Table 3. Target audience of the use cases.

Target audience	Values, n (%)	Study (reference)
Population or public health and health services research practitioners	53 (96)	Abusharekh et al, 2015 [67]; Afzal et al, 2011 [85]; Ali et al, 2016 [68]; Alonso et al, 2012 [92]; Antoniou et al, 2010 [93]; Baytas et al, 2016 [80]; Benis et al, 2017 [89]; Bryan et al, 2015 [64]; Byrd et al, 2016 [94]; Castronovo et al, 2009 [77]; Chen et al, 2016 [95]; Chorianopoulos et al, 2016 [96]; Dagliati et al, 2018 [66]; Deodhar et al, 2015 [65]; Garcia-Marti et al, 2017 [97]; Gligorijevi et al, 2017 [98]; Gotz et al, 2014 [76]; Guo et al, 2007 [69]; Haque et al, 2014 [99]; Hardisty et al, 2010 [100]; Huang et al, 2015 [101]; Hund et al, 2016 [90]; Ji et al, 2012 [102]; Ji et al, 2013 [81]; Jiang et al, 2016 [103]; Jinpon et al, 2017 [83]; Kaieski et al, 2016 [104]; Katsis et al, 2017 [105]; Kostkova et al, 2014 [75]; Kruzikas et al, 2014 [106]; Lavrac et al, 2007 [70]; Lu et al, 2017 [71]; Luo et al, 2016 [78]; Maciejewski et al, 2011 [79]; Marek et al, 2015 [108]; Mitrpanont et al, 2017 [109]; Mittelstadt et al, 2014 [110]; Ozkaynak et al, 2015 [111]; Park et al, 2018 [112]; Perer et al, 2015 [113]; Proulx et al, 2006 [114]; Shaban-Nejad et al, 2017 [84]; Soulakis et al, 2015 [115]; Tate et al, 2014 [87]; Tilahun et al, 2014 [88]; Toddenroth et al, 2014 [116]; Torres et al, 2012 [117]; Widanagamaachchi et al, 2017 [72]; Xing et al, 2010 [91]; Xu et al, 2013 [73]; Yan et al, 2013 [118]; Yu et al, 2017 [82]; Yu et al, 2018 [74]
Academics and data scientists	47 (85)	Abusharekh et al, 2015 [67]; Afzal et al, 2011 [85]; Antoniou et al, 2010 [93]; Baytas et al, 2016 [80]; Bryan et al, 2015 [64]; Byrd et al, 2016 [94]; Chorianopoulos et al, 2016 [96]; Dagliati et al, 2018 [66]; Deodhar et al, 2015 [65]; Garcia-Marti et al, 2017 [97]; Gligorijevi et al, 2017 [98]; Gotz et al, 2014 [76]; Guo et al, 2007 [69]; Haque et al, 2014 [99]; Hardisty et al, 2010 [100]; Huang et al, 2015 [101]; Hund et al, 2016 [90]; Ji et al, 2012 [102]; Ji et al, 2013 [81]; Jiang et al, 2016 [103]; Jinpon et al, 2017 [83]; Kaieski et al, 2016 [104]; Katsis et al, 2017 [105]; Kostkova et al, 2014 [75]; Kruzikas et al, 2014 [106]; Lavrac et al, 2007 [70]; Lu et al, 2017 [71]; Luo et al, 2016 [78]; Maciejewski et al, 2010 [107]; Maciejewski et al, 2011 [79]; Marek et al, 2015 [108]; Mitrpanont et al, 2017 [109]; Mittelstadt et al, 2014 [110]; Ozkaynak et al, 2015 [111]; Park et al, 2018 [112]; Perer et al, 2015 [113]; Proulx et al, 2006 [114]; Tate et al, 2014 [87]; Tilahun et al, 2014 [88]; Toddenroth et al, 2014 [116]; Torres et al, 2012 [117]; Widanagamaachchi et al, 2017 [72]; Xing et al, 2010 [91]; Xu et al, 2013 [73]; Yan et al, 2013 [118]; Yu et al, 2017 [82]; Yu et al, 2018 [74]
Clinicians	21 (38)	Abusharekh et al, 2015 [67]; Alonso et al, 2012 [92]; Antoniou et al, 2010 [93]; Baytas et al, 2016 [80]; Benis et al, 2017 [89]; Bryan et al, 2015 [64]; Chorianopoulos et al, 2016 [96]; Dagliati et al, 2018 [66]; Gotz et al, 2014 [76]; Haque et al, 2014 [99]; Huang et al, 2015 [101]; Hund et al, 2016 [90]; Lu et al, 2017 [71]; Mitrpanont et al, 2017 [109]; Mittelstadt et al, 2014 [110]; Ozkaynak et al, 2015 [111]; Perer et al, 2015 [113]; Soulakis et al, 2015 [115]; Toddenroth et al, 2014 [116]; Widanagamaachchi et al, 2017 [72]; Xu et al, 2013 [73]
Policy and decision makers	7 (13)	Ji et al, 2013 [81]; Kruzikas et al, 2014 [106]; Maciejewski et al, 2011 [79]; Mitrpanont et al, 2017 [109]; Tilahun et al, 2014 [88]; Torres et al, 2012 [117]; Yu et al, 2017 [82]
Consumers and public	5 (9)	Antunes de Mendonca et al, 2015 [86]; Ji et al, 2013 [81]; Kaieski et al, 2016 [104]; Maciejewski et al, 2011 [79]; Yu et al, 2017 [82]
Industry (software, pharmaceutical, and insurance)	3 (5)	Gotz et al, 2014 [76]; Perer et al, 2015 [113]; Yu et al, 2018 [74]

Terminology Related to Visualization and Analytics

We searched for the use of VA and its variations in the articles. Terms that indicated the use of VA included “visualization” (52/55, 95%), “visual analytics” (29/55, 53%), “analytics” (16/55, 29%), and “visual analytic approach” (8/55, 15%) as the employed method. The years when the term “visual analytics” was most commonly used were 2009 and 2017. As visualization was mentioned in the vast majority of the articles, alternative terms used for the analytic engine included data mining and ML techniques. Figure 4 displays the use of terms by year.

Figure 4. Use of terminology from January 01, 2005, to March 30, 2019.



Tool Name, Analytic Capability, and Goal

Related to the use of specific tools, articles that mentioned the name of the tools or base applications (34/55, 68%) are listed in Textbox 3.

In terms of analytic capability, these included tools primarily meant for descriptive analytics (52/55, 95%), exploratory analyses of complex data sets (23/55, 42%), and predictive analytics (13/55, 24%). There were no articles on prescriptive analytics. Among the overlap in the analytic capability of the tools, 3 studies (5%) mentioned descriptive and predictive analytic capabilities along with visual exploration of complex data sets [64-66]. There were 11 applications with both descriptive and predictive analytic capabilities [64-74].

We further categorized whether the application, tool, or method targeted population health and HSR decision support, knowledge discovery, or both, as its goal. There was an overlap in the application goals: decision support (44/55, 80%), knowledge discovery (35/55, 64%), or both goals (29/55, 53%).

Textbox 3. Name of the tool and base application (if provided).

Abusharekh et al, 2015 [67]

- H-Drive; information analytics based on R

Ali et al, 2016 [68]

- ID-Viewer

Alonso et al, 2012 [92]

- EPIPOI based on Matlab

Antoniou et al, 2010 [93]

- dAUTObase

Antunes de Mendonca et al, 2015 [86]

- On the basis of Triplify, SQL, PHP, and SPARQL EndPoint

Baytas et al, 2016 [80]

- PhenoTree

Benis et al, 2017 [89]

- DisEpi, R-based

Bryan et al, 2015 [64]

- EpiSimS

Chen et al, 2016 [95]

- SaTScan software

Chorianopoulos et al, 2016 [96]

- Flutrack.org

Dagliati et al, 2018 [66]

- MOSAIC dashboard; data mining using R and Matlab; JavaScript; HTML; Google Charts for GUI

Deodhar et al, 2015 [65]

- EpiCaster

Haque et al, 2014 [99]

- Microsoft SQL Server’s BI tool stack and ASP.NET

Hardisty et al, 2010 [100]

- LISTA-Viz

Hund et al, 2016 [90]

- Sub-VIS; based on D3.JS2

Ji et al, 2012 [102]

- ESMOS (Epidemic sentiment monitoring system)

Ji et al, 2013 [81]

- ESMOS (Epidemic sentiment monitoring system)

Jiang et al, 2016 [103]

- Health-Terrain

Jinpon et al, 2017 [83]

- Community well-being assessment system (CWBAS)

Kaieski et al, 2016 [104]

- Vis-Health

Kostkova et al, 2014 [75]

- medi+board

Lavrac et al, 2007 [70]

- MediMap

Lu et al, 2017 [71]

- Southampton breast cancer data system (SBCDS)

Luo et al, 2016 [78]

- GS-EpiViz

Maciejewski et al, 2011 [79]

- PanViz

Marek et al, 2015 [108]

- R with spacetime, gstat and plotKML; and Google Earth

Mitrpanont et al, 2017 [109]

- SAGE2

Ozkaynak et al, 2015 [111]

- EventFlow and Discrete Time Markov Chains

Perer et al, 2015 [113]

- Care pathway explorer

Proulx et al, 2006 [114]

- nSpace and GeoTime

Shaban-Nejad et al, 2017 [84]

- Population health record (PopHR)

Tate et al, 2014 [87]

- TrialViz

Yan et al, 2013 [118]

- ISS (syndromic surveillance system)

Yu et al, 2017 [82]

- Patient-provider geographic map

Yu et al, 2018 [74]

- Watson Analytics

Afzal et al, 2011 [85]; Byrd et al, 2016 [94]; Castronovo et al, 2009 [77]; Garcia-Marti et al, 2017 [97]; Gligorijevi et al, 2017 [98]; Gotz et al, 2014 [76]; Guo et al, 2007 [69]; Huang et al, 2015 [101]; Katsis et al, 2017 [105]; Kruzikas et al, 2014 [106]; Maciejewski et al, 2010 [107]; Mittelstadt et al, 2014 [110]; Park et al, 2018 [112]; Soulakis et al, 2015 [115]; Tilahun et al, 2014 [88]; Toddenroth et al, 2014 [116]; Torres et al, 2012 [117]; Widanagamaachchi et al, 2017 [72]; Xing et al, 2010 [91]; Xu et al, 2013 [73]

- Not mentioned

Multimedia Appendix 4 [64-118] details the analytic capability and goals of the application, indicating whether the analysis was carried out for knowledge discovery or decision support, whether the article was presented as a framework for VA, and whether the methodology itself followed one or more frameworks.

Framework Presented or Followed

A total of 24% (13/55) articles presented frameworks for VA methods, which we categorized into 7 types based on the major theories, applications, and functions that the study authors purported to use in their methods:

1. Data integration, monitoring, and management [67,68,71,75].
2. Combining querying, mining, and visualization for electronic medical records (EMRs) [76].
3. Disease mapping, hypotheses generation, clinical decision making, and knowledge discovery [66,77,78].
4. Simulation and modeling, including statistical analysis [64,79].
5. Phenotyping for a VA tool [80].
6. Social media VA [81].
7. Studying geographic variations in access to care [82].

A total of 29% (16/55) articles used a framework in their methods, which we broadly categorized into 6 types based on their application to the use case:

1. Studying access to care [83,84].
2. Analytics [78].
3. Application development [65,67,85,86].
4. Data quality, linkage, and flow [72,87,88].
5. Knowledge discovery [89].
6. Visualization [64,66,71,77,90].

Table 4 lists both kinds of frameworks and related references. The abovementioned categories are based on the objectives of the VA application, as mentioned by the authors in their studies.

Table 4. Articles proposing a framework and using frameworks for their visual analytics work with quoted references (if provided).

Study (reference)	Presents a framework	Uses one or more frameworks for VA[a] work
Abusharekh et al, 2015 [67]	Health data analytics framework incorporating data management, analytics, and visualization	Portal developed using Liferay and Vaadin frameworks
Afzal et al, 2011 [85]	N/A[b]	On the basis of the recommendations by Jankun-Kelly and Ma. [119]
Ali et al, 2016 [68]	Framework for data integration and analytics with various modules related to data acquisition, cleaning, parsing and analysis	N/A
Antunes de Mendonca et al, 2015 [86]	N/A	Resource development framework for queries, with SQL and others
Baytas et al, 2016 [80]	Phenotyping framework for a VA tool	N/A
Benis et al, 2017 [89]	N/A	Knowledge discovery in databases framework [120]
Bryan et al, 2015 [64]	Presents a framework for simulating and analyzing data. Visual engine also has a built-in statistical framework based on others	On the basis of the 3 frameworks [36-138]
Castronovo et al, 2009 [77]	Conceptual framework for dynamic mapping; hypotheses generation for disease seasonality	On the basis of the Harrower principles [139]
Dagliati et al, 2018 [66]	Presents a framework as a general model for chronic disease clinical decision support and knowledge discovery	Temporal abstraction [140]
Deodhar et al, 2015 [65]	N/A	Middleware based on the Model View Controller Framework
Gotz et al, 2014 [76]	Combines 3 components, such as visual query, pattern mining, and interactive vis components, in a single framework enabling an ad hoc event sequence analysis	N/A
Hund et al, 2016 [90]	N/A	Uses the detected subspaces of the OpenSubspace Framework and Visualization follows Shneiderman [141,142]
Ji et al, 2013 [81]	Framework considers several diseases; novel 2-step sentiment classification combining clue-based searching and ML methods to first filter out the nonpersonal; identifying all personal tweets; then distinguishing personal into negative and nonnegative sentiment tweets	N/A
Jinpon et al, 2017 [83]	N/A	Community Wellbeing Framework [143]
Kostkova et al, 2014 [75]	Framework depicts processes and components required for automated data monitoring across multiple real-time data channels [P Kostkova. A roadmap to integrated digital public health surveillance: the vision and the challenges. In Proceedings of the 22nd international conference on World Wide Web (WWW '13). 687-694., 2013]	N/A
Lu et al, 2017 [71]	Process-driven framework presented, with data, functional, and user layers	Lifelines framework sits within the University Hospital Southampton Clinical Data Environment as a model for the exploratory analysis of data
Luo et al, 2016 [78]	Presents a new framework for effective disease-control strategies, starting from identifying geo-social interaction patterns. Framework further used to structure the design of a VA tool with 3 components: reorderable matrix for geo-social mixing patterns, agent-based epidemic models, and combined visualization methods	Susceptible-Exposed-Infectious-Removed agent-based modeling
Maciejewski et al, 2011 [79]	The PanViz Visualization framework uses a mathematical epidemic model to calculate population dynamics and infection rate data	N/A
Shaban-Nejad et al, 2017 [84]	N/A	Semantic population health framework introduced in the tool by using type I evidence or causal knowledge to arrange health indicators along the lines of the determinants of health framework [144]
Tate et al, 2014 [87]	N/A	Data quality framework [145]
Tilahun et al, 2014 [88]	N/A	Silk Link Discovery Framework [146]
Widanagamaachchi et al, 2017 [72]	N/A	ViSUS framework for designing dataflow [147]

Study (reference)	Presents a framework	Uses one or more frameworks for VA[a] work
Yu et al, 2017 [82]	Introduces Visualization framework to aid health care policy makers and hospital administrators to visualize, identify, and optimize the geographic variations of access to care	N/A

[a] VA: visual analytics.

[b] N/A: not applicable.

Data Characteristics: Source, Use Cases, Structure, and Type

VA engines differ in their application, given their capability to process data from multiple data sets or various sources such as social media text data, administrative data, global repositories, and other internet sources. In the included studies, the data sources that were processed by the analytic engines varied, involving single data sources (32/55, 58%), multiple data sources (22/55, 40%), or both (6/55, 11%).

In use cases where multiple data sources were involved, there were overlaps within the categories of data sources: administrative or national survey data (17/55, 31%), EMR or EHR data (17/55, 31%), spatiotemporal data (16/55, 29%), web or social media data (8/55, 15%), and simulation data (6/55, 11%).

Articles focused on structured (40/55, 73%), unstructured (13/55, 24%), and semistructured data (5/55, 9%). The data sources were administrative data that included registry and national survey data (19/55, 35%), EMR or EHR data (17/55, 31%), spatiotemporal data (16/55, 29%), simulation data (6/55, 11%), and web or social media data (8/55, 15%). Multimedia Appendix 5 [64-118] details the source, type, and application to the use cases.

Analytic and Visualization Engines

From the articles, we gleaned information on the analytic engine, tools, and specific methods used, such as the algorithms for the analytic methods. The tool's analytic engine, its data processing, analysis, and subsequent data visualization varied greatly. In addition, details about the analytic and visualization engines have not been consistently reported.

We categorized the data for the type of problem that the application addressed by the major analytic techniques used for summarizing the results. There were 7 major categories: infectious disease modeling and surveillance (21/55, 38%); medical record pattern identification (20/55, 36%); population health monitoring (9/55, 16%); health system resource planning (2/55, 4%); and data manipulation, disease mapping, and sentiment analysis (1/55, 2%).

The analytic approaches undertaken included data querying (11/55, 20%); statistical modeling (11/55, 20%); clustering (9/55, 16%); natural language processing (NLP), pattern mining, classification, data mining, dimensionality reduction, predictive modeling, and other ML methods (4/55, 7%); and graph partitioning, neural networks, simulation-based predictions, and other statistical analyses (1/55, 2%).

The problems addressed and the analytic techniques used are summarized in Table 5. Multimedia Appendix 5 provides in-depth information on the data type, analytic and visual engines, and related techniques. Major tools employed for developing the applications included R-based tools (7/55, 13%); D3.JS (4/55, 7%); SQL (4/55, 7%); Java-based tools (3/55, 5%); Python-based tools, HTML 5, or Google Maps application programming interface (API; 2/55, 4%); and not reported (15/55, 27%). The 16 remaining articles mentioned the use of one of the following: Open Layers 3, OwlAPI, SaTScan, SQL and Google Maps API, IBM Watson Analytics, GeoViz Toolkit, Flutrack API, Weka, GeoTime, ESRI ArcMap, Excel2RDF and Sgvigler, C#, JFreeChart, MS Silverlight-based Pivot Viewer, Weka and Tableau, and Matlab.

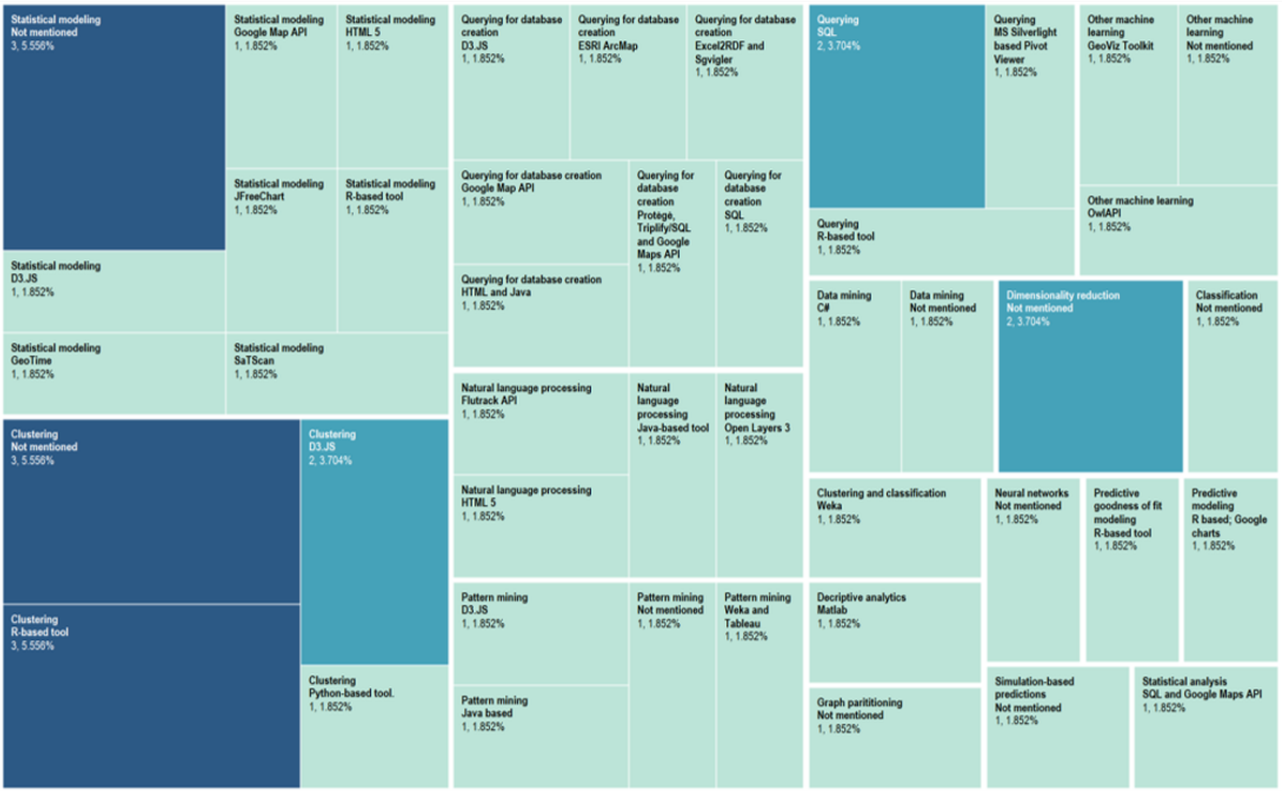
Table 5. Problem categories and major analytic methods.

Analytic method	Categories of problems with the number of articles mentioning the use of specific analytic methods							
	Data manipulation	Disease mapping	Health system resource planning	Infectious disease modeling and surveillance	Medical record pattern identification	Population health monitoring	Sentiment analysis	Total
Data querying	1	1	—[a]	3	5	1	—	11
Statistical modeling	—	—	1	8	2	—	—	11
Clustering	—	—	—	—	7	1	1	9
Natural language processing	—	—	—	3	1	—	—	4
Other machine learning	—	—	—	1	—	3	—	4
Pattern mining	—	—	—	1	3	—	—	4
Classification	—	—	1	—	—	1	—	2
Data mining	—	—	—	—	—	2	—	2
Dimensionality reduction	—	—	—	1	1	—	—	2
Predictive modeling	—	—	—	1	1	—	—	2
Graph partitioning	—	—	—	1	—	—	—	1
Neural networks	—	—	—	1	—	—	—	1
Simulation-based predictions	—	—	—	1	—	—	—	1
Statistical analysis	—	—	—	—	—	1	—	1
Total	1	1	2	21	20	9	1	55

[a] Null values.

The distribution of the tools used according to the analytic methods is illustrated in Figure 5. Among the most often used tools were R-based tools and packages, D3.JS, and Google Maps API. Almost all articles mentioned a different combination of tools that they had used for the VA application.

Figure 5. Analytic methods and proportional distribution of tools employed. API: application programming interface; MS: Microsoft; SQL: structured query language.



Similarly, various visualizations were used to represent the analysis of the data processed by the analytic engine. Visualizations were mostly interactive, with a dashboard presenting statistics or detailed information regarding specific populations or variables of interest. The major visualization types were as follows: maps (17/55, 31%), timelines (8/55, 15%), heat maps (6/55, 11%), choropleth maps (6/55, 11%), bubble charts (3/55, 5%), flow maps (2/55, 4%), and scatterplots (2/55, 4%). The remaining 12 applications presented one of the following visuals: spatial plots, history tree view, stacks and cards, line, bar, causal diagram, cards, stacked bar, population pyramid, circular tree, ranked trees, Sankey diagram, and relationship graph.

Domains of Health Care, Problem Category, and Related Analytic Methods

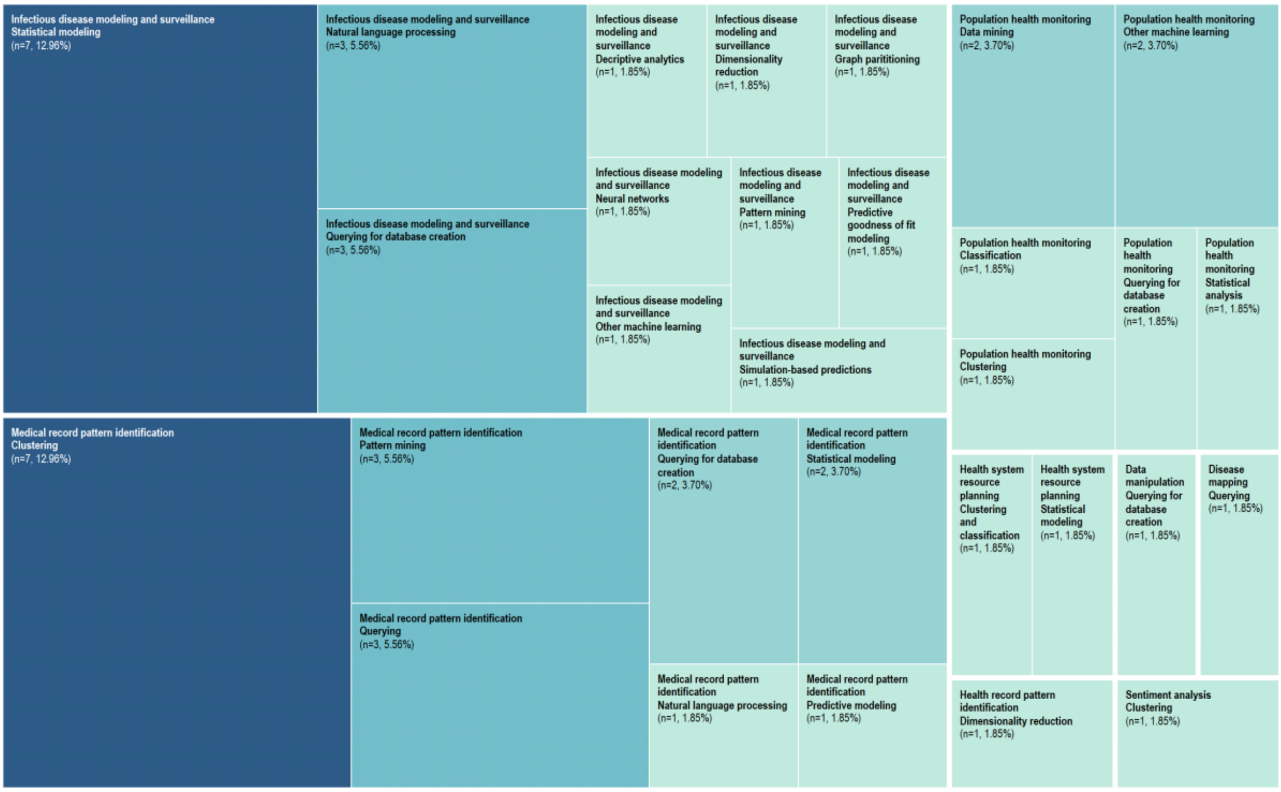
Of the articles, 98% (54/55) focused on population health, whereas 33% (18/55) focused on HSR. There was a considerable overlap, as 17 HSR articles had a population focus. Of the population health articles, 44% (24/55) were on clinical populations focusing on a condition or cluster of conditions and 31% (17/55) provided population demographics. Epidemic monitoring and modeling for certain conditions was the focus of 33% (18/55) studies, whereas 49% (27/55) were spatiotemporal health care articles.

Among the HSR articles, 27% (15/55) were on health service utilization, 18% (10/55) focused on access to care, and 4% (2/55) were related to health care costs. The details are provided in Multimedia Appendix 6 [64-118].

We further categorized the types of problems that the application addressed. The 4 major problem categories were infectious disease modeling and surveillance (21/55, 38%), medical record pattern identification (19/55, 35%), population health monitoring (9/55, 16%), and health system resource planning (2/55, 4%). One use case was for data manipulation, disease mapping, health record pattern identification, and sentiment analysis.

Figure 6 details the relative distribution of the analytic methods used for the categories of problems. The color-coded tree map reveals clustering and statistical modeling as the major choice for medical record pattern identification and infectious disease modeling and surveillance, both methods comprising 13% (7/55) of all use cases. The second most common methods included NLP, querying for database creation, pattern mining, and data querying, each comprising 5% (3/55) of all methods. Other varied methods are shown in the figure to reveal the overall trends found in the use of methods according to the problem addressed by the application.

Figure 6. Problem categories and proportional distribution of analytic methods used.



Availability of the Tool, Uptake, and Co-Design

A total of 21 VA tools were in use at the time of publication of the original article, whereas others were either not available or prototypes. Moreover, 7 dashboards or tools were accessible for public use, whereas 13 tools were either developed using free, open source tools such as R or Weka or their source code was provided. Furthermore, 38% (21/55) articles did not mention the tools used to develop the application or the base application.

Relevant to co-design or evaluation of the tool, 10 articles involved domain experts, multidisciplinary teams, or user evaluations for the development and improvement of the VA application, whereas other articles did not mention this aspect. Relevant details have been captured in Multimedia Appendix 7 [64-118].

Innovation and Limitations of VA Applications

All applications offered an innovative edge over others at the time of their publication. These mostly pertained to the analytic engine and techniques such as better workflow, automation, development of a framework, and use of advanced techniques such as ML. Similarly, the limitations of the applications were provided in varying detail, with 29% (16/55) articles not mentioning any limitations, as shown in Multimedia Appendix 7.

Discussion

Significance of the Review

The aim of this scoping review is to review the literature on VA methods, specifically their application to the fields of population health and HSR. Given the large variety, heterogeneity, and velocity of data sources, public health data belong to the category of big data [50], which are increasingly being generated and made available from administrative, EMR, and EHR sources. Examples of large population-level repositories include the United Kingdom’s Clinical Practice Research Datalink database, the largest collection of anonymized primary care patient records [87]; the Canadian administrative health data sets [121]; and the US National Health and Nutrition Examination Survey, collected since the 1970s [91].

Our scoping review summarizes VA methods applied to use cases in population health and HSR. As a multidisciplinary team, we presented the results from multiple perspectives, including those of data scientists, population health and HSR practitioners, and policy and decision makers. This is important in the wake of the COVID-19 pandemic, where multiple VA products for pandemic monitoring have surfaced for guiding the pandemic response [122,123].

We discuss the implications and contributions of this review for researchers and practitioners in the related health care areas of public and population health and HSR, expanding on aspects of specific import. We further offer targeted recommendations for defining, reporting, and leveraging the potential of VA methods and applications.

Reporting Checklist for VA Applications

The field of visualization and analytics is extremely broad, with various applications in different health care and other sectors. We had to rely mainly on iterative screening to filter out articles that were not relevant to the study objectives, for example, articles without use cases including usability studies, evaluations, human-computer interaction, and GIS studies.

There is a need for better reporting on the details of the applications for reproducibility and transparency. This specifically relates to the tool's capability, application beyond the use case, target audience, study objectives, and study settings. In many articles, we found the statistical and analytic methods lacking in detail, in particular on the tools used for the analysis, the algorithms tested and applied, and the reasons for choosing one particular analysis over the other.

Similarly, some articles from proprietary or prototypical tools did not offer any detail on the analytic engine, while only discussing the functional aspects of the application. Many articles did not elaborate on how visualization presentations should be interpreted. Most articles did not provide reasons and processes for the selection of the visualization, its strength over others, and how the interactive functions could offer more insight.

Such details would help situate the literature and resultantly be useful for better reproducibility, development, and adaptability of prototypical and established methods to different scenarios.

Towards this goal, we developed a standard reporting checklist (Multimedia Appendix 8) for reporting VA methods, particularly visual and analytics engines, as is the practice for reporting research methods such as statistical techniques [124], and qualitative and quantitative studies [125,126]. As mentioned previously, 4 recent systematic reviews covered areas of VA applications in health [8,17-19]. Although these reviews offered excellent summaries from different areas of health care and informatics, we found that there was no reporting standard followed, indicating the need for such a checklist. We further sought recent literature on COVID-19–related VA products. One of the most known COVID-19–related products is the web-based dashboard for country-level data by Johns Hopkins University [123]. Although the experts involved have not yet published a paper detailing the methods for its development, a high-level correspondence article was published in the reputed journal *Lancet* [123]. In this article, the authors cite the issues and process of developing a data stream for the dashboard. In contrast, the authors of another VA product mapping the COVID-19–related mobility pattern changes in US counties detail the methods, features of the web-based platform, data sources, system design, and insights from the results in their publication [122].

On the basis of the findings from these papers and those included in the scoping review, we have proposed a checklist for reporting VA applications (Multimedia Appendix 8) to fulfill the need for standard reporting aimed at optimizing productivity from research efforts [3].

Proposed Definition for VA in Health Care

We adhered to the definition for VA applications in health care by Ola and Sediq [50] for which both an analytic and visual engine must be included. However, we found that despite reporting analytic techniques, including an analytic engine, many articles did not state it as such. Although "visualization" as a term was mentioned in all articles, analytic techniques were not mainly classified as analytic engines. This could be due to the different use of language and understanding within the data science communities of practice. Hence, VA as a term with a technical definition does not seem firmly established, at least in the health care literature. This can also be seen in various authors' work where they borrow from the original definition of VA by Thomas and Cook [2] being "*the science of analytical reasoning facilitated by interactive visual interfaces.*"

Thomas and Cook [2] define VA as "the science of analytical reasoning facilitated by interactive visual interfaces," whereas Keim et al [3] extended the concept to "automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets." Borrowing from the seminal works of Ola and Sedig [50], Keim et al [3], and Thomas and Cook [2], we recommend using the following adapted definition of VA, especially in areas related to public and population health: "an approach, method or application for analytic reasoning, exploration, knowledge discovery, and sense making of complex data, through the use of one or more interactive visual interfaces, employing analytic and visual engines." In our definition, we keep the aim of the VA technique at the fore to provide context to the method, while expanding on the limited concept of VA to *computational tools* [3,50]. We emphasize the analytic and visual engines to help delineate the methods from other fields, such as visualization. We also emphasize it as it helps to define and report the methods better, for which we included a checklist for reporting (Multimedia Appendix 8).

VA Methods, Frameworks, and Tools

We followed a broad definition of frameworks to summarize the VA methods in developing the applications. Although presenting the detailed findings from these frameworks is beyond the scope of this review, we broadly categorized their types, as it can be valuable to learn from the conceptual and theoretical bases of this innovative method. Studying both types of frameworks helps situate the methods for adaptation by researchers and practitioners. Among the variety of VA frameworks presented, most were related to disease mapping and for knowledge discovery and hypothesis generation [66,77,78]. This is consistent with the findings of the goals and analytic capabilities of the tools that we summarize. Although there is a trend toward the application of ML methods to EMR data sets, we found 1 framework for mining and visualizing trends and patterns from these data sets [76].

The majority of the applications were prototypes, with only 5 in use at the time of publication. In part, this may be due to publication bias of newer VA techniques. Studies using proprietary tools known for their visualization engines, such as Tableau [148], Qlik [149], and Power BI [150], were exceptionally uncommon in the articles that were reviewed. Hence, future research may seek to survey the population health and HSR practitioners to gain an understanding of the VA tools that are part of their daily decision-making processes and reasons for not publishing their findings and experiences.

Settings, Target Audience, and Co-Design Initiatives

As we limited our search to English language articles, the use cases from the included studies were mostly from the United

States, with fewer studies from Europe or other countries. Most of the authors and use cases were from the same country, aside from one exception, an author from Canada working on a use case from the United States [91].

As expected, population health and HSR practitioners were the most common intended target audience, followed by academic researchers and clinicians. Policy and decision makers as well as the general public were not the main target audiences. Although VA is related to data visualization and is being increasingly employed to convey insights from the data, we contend that the use of VA is still in the developmental stages. This corresponds to our finding that most applications were prototypes.

There were 5 studies that were aimed at consumers, whereas 7 studies targeted policy and decision makers. However, as participatory approaches are being emphasized for better uptake and development of creative solutions [127], a concern related to co-designing applications was the lack of involvement of stakeholders, such as decision makers, and patient groups. It is important to note that the lack of participatory co-design approaches in developing applications could be one reason for the overall finding of slow uptake of these methods in population health and HSR.

Trends and Potential for VA Applications

As the results show, the use of VA varies greatly. In addition, due to inconsistent reporting of the settings and target audiences in the included papers, we made a calculated judgment on the trends in the use of these techniques. As most of the studies were conducted in academic circles, we infer that these methods are still in development in the population health and HSR communities of practice. Hence, the uptake of these methods has been slow in these interrelated areas of health care. This is not unexpected, as the field has been termed nascent, while the application of newer techniques in public health has been rather delayed [50].

As most tools focused on descriptive analytics, with about half aimed at visual exploration of complex data sets, the trend in the use of these methods toward knowledge discovery and decision support is notable. This could be due to the availability of increasing and expanded data sets from EMR systems. For ML, clustering, classification, and NLP are methods of choice for structured and text-based data sets. Many population health applications are related to mapping, spatiotemporal distribution, and modeling for diseases and disease control. In HSR, few articles dealt with issues of access, utilization, and costs of services.

Most problems addressed related to infectious disease epidemiology, with clustering and statistical modeling being the most commonly used analytic methods. The articles mentioned a unique tool, a combination of tools, or did not mention the tool or base application, which made it difficult to summarize the types of tools used. However, as shown in Figure 5, R-based tools, Google Maps API and D3.JS, as well as a variety of other tools were used for the VA applications.

In addition, there is added value in using VA to obtain and combine multiple data sources to construct a fuller picture toward the question of inquiry. As our results show, the analytic engine in most use cases combines multiple data sources, such as EMRs, to social media sources. As Keim et al [3] point out, VA can contribute to solving various complex problems in sectors including engineering, financial analysis, environment and climate change, and socio-economic conditions. Socio-economic considerations in health, known as the social determinants of health, are being increasingly researched in the context of accessibility, health, and overall quality of life of populations [128]. In addition, VA has the potential to address the varied shared application problems in health and related sectors at an abstract level [3].

Learning Health Systems and COVID-19–Related VA Products

Learning health systems are geared toward continuous evidence-based quality improvement [129]. There are multiple challenges in building such systems that generate knowledge and insights on proposed improvements [130]. In the wider context, this review allows fellow researchers, practitioners, and decision makers to appreciate the potential presented by VA techniques in meeting challenges in operationalizing and building automated data-driven learning health systems [131]. VA techniques have the ability for sense making and leveraging big data from multiple sources to operationalize such learning health systems [33,132].

As has been the case in the last few months of the COVID-19 crisis, a plethora of VA products have surfaced, aimed at clinical practitioners, population health and health service researchers, policy makers, and the general public [122,123,133]. Such VA products are being increasingly sought for epidemiologic surveillance, monitoring, and planning of health services, in addition to apprising the public on the magnitude of the pandemic. It will be especially useful for research replicability and transparency to describe the development and features of such products in sufficient detail, toward which we presented a reporting checklist (Multimedia Appendix 8), for aspects that we found to be important in reporting methods and functionality of an application. We are confident that this will serve novice and expert researchers alike as a reminder to showcase the depth and breadth of their efforts in developing a unique application.

Limitations

We based our inquiry of VA methods on information from peer-reviewed journal articles and full conference papers. We did not include book chapters, theses, short papers, editorials, non–peer-reviewed reports, conference abstracts, and live websites using VA techniques. We limited our review to the year 2005 onward, and we did not explore subject-specific databases from mathematics, geography, and computer sciences. We sought to limit our findings to proposed or established methods that have been either published or presented and applied to actual use cases. We included full conference papers in the review, but many conferences do not publish proceedings, such as the annual Tableau conference and the Health Analytics Summit. Use cases discussed at these meetings mostly involve front-end proprietary tools. Hence, the complete spectrum of the use of such tools could not be covered in this review. However, we followed the highest methodological standards for conducting systematic reviews. This included developing a multidisciplinary team of health researchers and data scientists, following established review frameworks with at least two independent reviewers at each step, and being guided by a dedicated information specialist. Our search strategy was developed over multiple iterations and was peer reviewed using the PRESS guidelines [134] by an independent third-party information specialist, whereas we published the review protocol in advance [21].

Conclusions

VA as an innovative field holds great potential in yielding insights from big health care data, especially in the related fields of population health and HSR. This is especially relevant in the backdrop of the COVID-19 pandemic, where multiple VA products have taken center stage.

This scoping review provides a foundational understanding of the current landscape on the application of VA methods in areas of population health and HSR. We present the major VA tools, techniques, and frameworks since 2005 published in peer-reviewed papers. VA is an innovative, rapidly expanding field with its roots in many disciplines, and it is being used to build learning health systems for improving patient care, increasing access to services, controlling costs, and appropriately allocating resources [33]. It is expected that the next generation of EMR systems will leverage advanced analytics to meet the needs of diverse audiences [135]. Such systems are aimed at harmonizing patient records; creating a seamless picture of access to care at primary, secondary, and tertiary levels; and managing disease outbreaks at the population level. We also present an expanded definition for VA applications in health care, with a reporting checklist to help researchers provide solutions for greater replicability.

Acknowledgments

The review is part of the doctoral work under JC's Canadian Institutes of Health Research (CIHR) funded Health System Impact Fellowship 2018-19 and 2019-20. JC, IB, MM, MI, AK, JM, and LT are Health System Impact Fellows cofunded by the Canadian Institutes of Health Research and their host organizations. The authors' host organizations are mentioned with their affiliation. SJ holds the Toronto Rehabilitation Institute Chair at the University of Toronto. The authors wish to thank JC's Thesis Committee Members, Professor Cathy Craven, Professor Sara Guilcher and Professor Christopher Collins for review and critical feedback on the manuscript. Aldo Barrera, MSc contributed significantly to the Tableau dashboards to present the results in an interactive visual format.

Authors' Contributions

All authors contributed significantly to the conceptualization and reporting of the review. JC, IB, JM, MI, MM, SM, SJ, and TJ conceptualized the review, screened articles, contributed to the abstraction, and reviewed the manuscript. JB constructed, translated, and executed search strategies. The first 4 authors, JC, JM, IB, and LT, mainly wrote and revised the manuscript in consultation with others. SA, MB, SH, and MI contributed to the abstraction and validation of data and preparation of the manuscript. All authors critically reviewed and revised the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) Checklist for reporting scoping reviews.

[PDF File (Adobe PDF File), 605 KB - jmir_v22i12e17892_app1.pdf]

Multimedia Appendix 2

MEDLINE search strategy.

[DOCX File , 47 KB - jmir_v22i12e17892_app2.docx]

Multimedia Appendix 3

Study characteristics, problem analyzed, settings, and target audience.

[PDF File (Adobe PDF File), 148 KB - jmir_v22i12e17892_app3.pdf]

Multimedia Appendix 4

Analytic capability and goals.

[PDF File (Adobe PDF File), 131 KB - jmir_v22i12e17892_app4.pdf]

Multimedia Appendix 5

Data types and analytic and visual engines.
[PDF File (Adobe PDF File), 238 KB - jmir_v22i12e17892_app5.pdf]

Multimedia Appendix 6
Domains of health care.
[PDF File (Adobe PDF File), 87 KB - jmir_v22i12e17892_app6.pdf]

Multimedia Appendix 7
Key findings, impact, innovation, availability, and limitations.
[PDF File (Adobe PDF File), 145 KB - jmir_v22i12e17892_app7.pdf]

Multimedia Appendix 8
Checklist for reporting Visual Analytic applications in population health and health services research.
[DOCX File , 26 KB - jmir_v22i12e17892_app8.docx]

References

1. Wong PC, Thomas J. Visual analytics. In: IEEE Computer Graphics and Applications. Vienna, Austria: IEEE; 2004.
2. Thomas K, Cook K. Illuminating the Path: the R&d Agenda for Visual Analytics. Richland, Washington: United States, Department of Homeland Security; 2011.
3. Keim D, Andrienko G, Fekete J, Carsten G, Melan G, Keim D. Visual analytics: definition, process and challenges. InfoVis 2008:154-175. [doi: 10.1007/978-3-540-70956-5_7]
4. Keim D, Zhang L. Solving Problems With Visual Analytics. In: Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies. 2011 Presented at: KMKT'11; 2011; Graz, Austria URL: https://www.sciencedirect.com/science/article/pii/S1877050911007009
5. Andrienko G, Andrienko N, Demsar U, Dransch D, Dykes J, Fabrikant SI, et al. Space, time and visual analytics. Int J Geog Inf Sci 2010 Oct 11;24(10):1577-1600. [doi: 10.1080/13658816.2010.508043]
6. Caban JJ, Gotz D. Visual analytics in healthcare--opportunities and research challenges. J Am Med Inform Assoc 2015 Mar;22(2):260-262. [doi: 10.1093/jamia/ocv006] [Medline: 25814539]
7. Simpao AF, Ahumada LM, Gálvez JA, Rehman MA. A review of analytics and clinical informatics in health care. J Med Syst 2014 Apr;38(4):45. [doi: 10.1007/s10916-014-0045-x] [Medline: 24696396]
8. Chung Y, Bagheri N, Salinas-Perez JA, Smurthwaite K, Walsh E, Furst M, et al. Role of visual analytics in supporting mental healthcare systems research and policy: a systematic scoping review. Int J Inf Manag 2020 Feb;50:17-27 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2019.04.012]
9. Kindig D, Stoddart G. What is population health? Am J Public Health 2003 Mar;93(3):380-383. [doi: 10.2105/ajph.93.3.380] [Medline: 12604476]
10. Kindig D. Understanding population health terminology. Milbank Q 2007;85(1):139-161 [FREE Full text] [doi: 10.1111/j.1468-0009.2007.00479.x] [Medline: 17319809]
11. Health Services Research. Canadian Institutes of Health Research. 2019. URL: http://www.cihr-irsc.gc.ca/e/48809.html [accessed 2019-03-13]
12. Lohr KN, Steinwachs DM. Health services research: an evolving definition of the field. Health Serv Res 2002 Feb;37(1):15-17. [doi: 10.1111/1475-6773.01020]
13. Martin-Sanchez F, Verspoor K. Big data in medicine is driving big changes. Yearb Med Inform 2018 Mar 5;23(01):14-20 [FREE Full text] [doi: 10.15265/iy-2014-0020]
14. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
15. White SE. A review of big data in health care: challenges and opportunities. OAB 2014 Oct 31;6:13-18. [doi: 10.2147/oab.s50519]
16. Simpao AF, Ahumada LM, Rehman MA. Big data and visual analytics in anaesthesia and health care. Br J Anaesth 2015 Sep;115(3):350-356 [FREE Full text] [doi: 10.1093/bja/aeu552] [Medline: 25627395]
17. Islam MS, Hasan MM, Wang X, Germack HD, Noor-E-Alam M. A systematic review on healthcare analytics: application and theoretical perspective of data mining. Healthcare (Basel) 2018 May 23;6(2):- [FREE Full text] [doi: 10.3390/healthcare6020054] [Medline: 29882866]
18. West VL, Borland D, Hammond WE. Innovative information visualization of electronic health record data: a systematic review. J Am Med Inform Assoc 2015 Mar;22(2):330-339 [FREE Full text] [doi: 10.1136/amiajnl-2014-002955] [Medline: 25336597]
19. Wu D, Chen A, Manning J, Levy-Fix G, Backonja U, Borland D, et al. Evaluating visual analytics for health informatics applications: a systematic review from the American Medical Informatics Association Visual Analytics Working Group

Task Force on Evaluation. J Am Med Inform Assoc 2019 Apr 1;26(4):314-323. [doi: 10.1093/jamia/ocy190] [Medline: 30840080]

20. Braunstein ML. Practitioner's Guide to Health Informatics. Switzerland: Springer, Cham; 2015.
21. Chishtie JA, Babineau J, Bielska IA, Cepoiu-Martin M, Irvine M, Koval A, et al. Visual analytic tools and techniques in population health and health services research: protocol for a scoping review. JMIR Res Protoc 2019 Oct 28;8(10):e14019 [FREE Full text] [doi: 10.2196/14019] [Medline: 31661081]
22. Peters MD, Godfrey CM, McInerney P, Soares CB, Khalil H, Parker D. Joanna Briggs Institute Reviewers' Manual: 2015 Edition. Australia: The Joanna Briggs Institute; 2015.
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Method 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
24. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
25. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-SCR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
27. Cleo G, Scott AM, Islam F, Julien B, Beller E. Usability and acceptability of four systematic review automation software packages: a mixed method design. Syst Rev 2019 Jun 20;8(1):145 [FREE Full text] [doi: 10.1186/s13643-019-1069-6] [Medline: 31221212]
28. Shneiderman B, Plaisant C, Hesse BW. Improving healthcare with interactive visualization. Computer 2013 May;46(5):58-66 [FREE Full text] [doi: 10.1109/mc.2013.38]
29. Viju Raghupathi WR. An overview of health analytics. J Health Med Informat 2013 Oct 7;4(3):2 [FREE Full text] [doi: 10.4172/2157-7420.1000132]
30. Lin R, Ye Z, Wang H, Wu B. Chronic diseases and health monitoring big data: a survey. IEEE Rev Biomed Eng 2018;11:275-288. [doi: 10.1109/rbme.2018.2829704]
31. Liao H, Tang M, Luo L, Li C, Chiclana F, Zeng X. A bibliometric analysis and visualization of medical big data research. Sustainability 2018 Jan 11;10(2):166. [doi: 10.3390/su10010166]
32. Khalifa M. Health analytics types, functions and levels: a review of literature. Stud Health Technol Inform 2018;251:137-140. [Medline: 29968621]
33. Fihn SD, Francis J, Clancy C, Nielson C, Nelson K, Rumsfeld J, et al. Insights from advanced analytics at the veterans health administration. Health Aff (Millwood) 2014 Jul;33(7):1203-1211. [doi: 10.1377/hlthaff.2014.0054] [Medline: 25006147]
34. Docherty SL, Vorderstrasse A, Brandon D, Johnson C. Visualization of multidimensional data in nursing science. West J Nurs Res 2017 Jan;39(1):112-126 [FREE Full text] [doi: 10.1177/0193945916672448] [Medline: 27756850]
35. Backonja U, Chi N, Choi Y, Hall AK, Le T, Kang Y, et al. Visualization approaches to support healthy aging: a systematic review. J Innov Health Inform 2016 Oct 10;23(3):860 [FREE Full text] [doi: 10.14236/jhi.v23i3.860] [Medline: 28059694]
36. What is Public Health? CDC Foundation. 2020. URL: https://www.cdcfoundation.org/what-public-health [accessed 2020-11-14]
37. Niedźwiedzka B, Czabanowska K, Śmietana R. Controlled vocabulary in public health. An overview of the achievements to date. J Public Health 2008 Jun 25;17(1):55-59. [doi: 10.1007/s10389-008-0212-y]
38. Population Health. National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/mesh/2023448
39. UK: The Health Agency, Health Development List. National Public Health Language. 2007. URL: https://en.wikipedia.org/wiki/Health_care_in_the_United_Kingdom [accessed 2020-11-20]
40. Etches V, Frank J, di Ruggiero E, Manuel D. Measuring population health: a review of indicators. Annu Rev Public Health 2006;27:29-55. [doi: 10.1146/annurev.publhealth.27.021405.102141] [Medline: 16533108]
41. Cohen D, Huynh T, Sebold A, Harvey J, Neudorf C, Brown A. The population health approach: a qualitative study of conceptual and operational definitions for leaders in Canadian healthcare. SAGE Open Med 2014;2:2050312114522618 [FREE Full text] [doi: 10.1177/2050312114522618] [Medline: 26770704]
42. Fone D, Hollinghurst S, Temple M, Round A, Lester N, Weightman A, et al. Systematic review of the use and value of computer simulation modelling in population health and health care delivery. J Public Health Med 2003 Dec;25(4):325-335. [doi: 10.1093/pubmed/fdg075] [Medline: 14747592]
43. Singh SR. Public health spending and population health: a systematic review. Am J Prev Med 2014 Nov;47(5):634-640. [doi: 10.1016/j.amepre.2014.05.017] [Medline: 25084684]
44. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. Annu Rev Public Health 2016;37:61-81 [FREE Full text] [doi: 10.1146/annurev-publhealth-032315-021353] [Medline: 26667605]

45. Harris JK, Beatty KE, Barbero C, Howard AF, Cheskin RA, Shapiro RM, et al. Methods in public health services and systems research: a systematic review. Am J Prev Med 2012 May;42(5 Suppl 1):S42-S57. [doi: 10.1016/j.amepre.2012.01.028] [Medline: 22502925]
46. Thompson DS, Fazio X, Kustra E, Patrick L, Stanley D. Scoping review of complexity theory in health services research. BMC Health Serv Res 2016 Mar 12;16:87 [FREE Full text] [doi: 10.1186/s12913-016-1343-4] [Medline: 26968157]
47. Gulley SP, Rasch EK, Bethell CD, Carle AC, Druss BG, Houtrow AJ, et al. At the intersection of chronic disease, disability and health services research: a scoping literature review. Disabil Health J 2018 Apr;11(2):192-203 [FREE Full text] [doi: 10.1016/j.dhjo.2017.12.012] [Medline: 29396271]
48. Rowland M, Peterson-Besse J, Dobbertin K, Walsh ES, Horner-Johnson W, Expert Panel on Disability and Health Disparities. Health outcome disparities among subgroups of people with disabilities: a scoping review. Disabil Health J 2014 Apr;7(2):136-150. [doi: 10.1016/j.dhjo.2013.09.003] [Medline: 24680042]
49. Finding and Retrieving HSR: Tools and Databases. US National Library of Medicine. 2018. URL: https://wayback.archive-it.org/org-350/20180515152423/https://www.nlm.nih.gov/nichsr/db.html [accessed 2019-03-13]
50. Ola O, Sedig K. The challenge of big data in public health: an opportunity for visual analytics. Online J Public Health Inform 2014;5(3):223 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4933] [Medline: 24678376]
51. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
52. Pike WA, Stasko J, Chang R, O'Connell TA. The science of interaction. Inf Vis 2009 Jan 24;8(4):263-274. [doi: 10.1057/ivs.2009.22]
53. Basole R, Braunstein M, Kumar V, Park H, Kahng M, Chau DH, et al. Understanding variations in pediatric asthma care processes in the emergency department using visual analytics. J Am Med Inform Assoc 2015 Mar;22(2):318-323. [doi: 10.1093/jamia/ocu016] [Medline: 25656514]
54. Basole R, Park H, Gupta M, Braunstein M, Chau D, Thompson M. A Visual Analytics Approach to Understanding Care Process Variation and Conformance. In: Proc 2015 Workshop Visual Analytical Healthcare. 2015 Presented at: VAHC'15; October 3-6, 2015; Chicago Illinois URL: http://europepmc.org/abstract/MED/29177250 [doi: 10.1145/2836034.2836040]
55. Falster MO, Jorm LR, Leyland AH. Visualising linked health data to explore health events around preventable hospitalisations in NSW Australia. BMJ Open 2016 Sep 7;6(9):e012031 [FREE Full text] [doi: 10.1136/bmjopen-2016-012031] [Medline: 27604087]
56. Hosseinpoor AR, Schlotheuber A, Nambiar D, Ross Z. Health equity assessment toolkit plus (HEAT Plus): software for exploring and comparing health inequalities using uploaded datasets. Glob Health Action 2018;11(sup1):1440783 [FREE Full text] [doi: 10.1080/16549716.2018.1440783] [Medline: 29974823]
57. Huang C, Syed-Abdul S, Jian W, Iqbal U, Nguyen P, Lee P, et al. A novel tool for visualizing chronic kidney disease associated polymorbidity: a 13-year cohort study in Taiwan. J Am Med Inform Assoc 2015 Mar;22(2):290-298. [doi: 10.1093/jamia/ocu044] [Medline: 25814540]
58. Martinez R, Ordunez P, Soliz PN, Ballesteros MF. Data visualisation in surveillance for injury prevention and control: conceptual bases and case studies. Inj Prev 2016 Apr;22(Suppl 1):i27-i33 [FREE Full text] [doi: 10.1136/injuryprev-2015-041812] [Medline: 26728006]
59. Ratwani RM, Fong A. 'Connecting the dots': leveraging visual analytics to make sense of patient safety event reports. J Am Med Inform Assoc 2015 Mar;22(2):312-317. [doi: 10.1136/amiajnl-2014-002963] [Medline: 25336592]
60. Grant C, Osanloo A. Understanding, selecting, and integrating a theoretical framework in dissertation research: creating the blueprint for your 'house'. Admin Int J 2014;4(2):12-26. [doi: 10.5929/2014.4.2.9]
61. Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. Annu Rev Public Health 2007;28:413-433. [doi: 10.1146/annurev.publhealth.28.021406.144145] [Medline: 17150029]
62. Ward ME, de Brún A, Beirne D, Conway C, Cunningham U, English A, et al. Using co-design to develop a collective leadership intervention for healthcare teams to improve safety culture. Int J Environ Res Public Health 2018 Jun 5;15(6):- [FREE Full text] [doi: 10.3390/ijerph15061182] [Medline: 29874883]
63. Jawad Chishtie. Tableau Public: Free Data Visualization Software. 2019. URL: https://public.tableau.com/profile/jawad.chishtie#!/ [accessed 2020-11-24]
64. Bryan C, Wu X, Mniszewski S, Ma K. Integrating Predictive Analytics Into a Spatiotemporal Epidemic Simulation. In: IEEE Conference on Visual Analytics Science and Technology. 2015 Presented at: VAST'15; October 25-30, 2015; Chicago, IL, USA. [doi: 10.1109/VAST.2015.7347626]
65. Deodhar S, Bisset K, Chen J, Barrett C, Wilson M, Marathe M. EpiCaster: an integrated web application for situation assessment and forecasting of global epidemics. ACM BCB 2015 Sep;2015:156-165 [FREE Full text] [doi: 10.1145/2808719.2808735] [Medline: 27796009]
66. Dagliati A, Sacchi L, Tibollo V, Cogni G, Teliti M, Martinez-Millana A, et al. A dashboard-based system for supporting diabetes care. J Am Med Inform Assoc 2018 May 1;25(5):538-547. [doi: 10.1093/jamia/ocx159] [Medline: 29409033]
67. Abusharekh A, Stewart S, Hashemian N, Abidi S. H-DRIVE: A Big Health Data Analytics Platform for Evidence-Informed Decision Making. In: IEEE International Congress on Big Data. 2015 Presented at: CBD'15; June 27-July 2, 2015; New York, USA. [doi: 10.1109/BigDataCongress.2015.68]

68. Ali M, Ahsan Z, Amin M, Latif S, Ayyaz A, Ayyaz M. ID-Viewer: a visual analytics architecture for infectious diseases surveillance and response management in Pakistan. Public Health 2016 May;134:72-85. [doi: 10.1016/j.puhe.2016.01.006] [Medline: 26880489]
69. Guo D. Visual analytics of spatial interaction patterns for pandemic decision support. Int J Geog Inform Sci 2007 Sep;21(8):859-877. [doi: 10.1080/13658810701349037]
70. Lavrac N, Bohanec M, Pur A, Cestnik B, Debeljak M, Kobler A. Data mining and visualization for decision support and modeling of public health-care resources. J Biomed Inform 2007 Aug;40(4):438-447 [FREE Full text] [doi: 10.1016/j.jbi.2006.10.003] [Medline: 17157076]
71. Lu J, Hales A, Rew D. Modelling of Cancer Patient Records: A Structured Approach to Data Mining and Visual Analytics. Lyon, France: Springer; 2017.
72. Widanagamaachchi W, Livnat Y, Bremer P, Duvall S, Pascucci V. Interactive visualization and exploration of patient progression in a hospital setting. AMIA Annu Symp Proc 2017;2017:1773-1782 [FREE Full text] [Medline: 29854248]
73. Xu S, Jewell B, Steed C, Schryver J. A New Collaborative Tool for Visually Understanding National Health Indicators. In: Proceedings of the International Conference on Applied Human Factors and Ergonomics 2. 2013 Presented at: AHFE'13; January 2-5, 2012; San Francisco, US p. 91-100. [doi: 10.1201/b12318-14]
74. Yu N, Zheng M, Andrade X, Patane R. Visual Analysis for Exploring the Relation Between Air Pollution, Environmental Factors and Respiratory Diseases. In: 11th EAI International Conference on Mobile Multimedia Communications. 2018 Presented at: EAI'18; July 7-9, 2018; Qingdao, China. [doi: 10.4108/eai.21-6-2018.2276626]
75. Kostkova P, Garbin S, Moser J, Pan W. Integration and Visualization Public Health Dashboard: the Medi+board Pilot Project. In: International World Wide Web Conference. 2014 Presented at: WWW'14; April 12-14, 2014; Seoul Korea, Japan. [doi: 10.1145/2567948.2579276]
76. Gotz D, Wang F, Perer A. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. J Biomed Inform 2014 Apr;48:148-159 [FREE Full text] [doi: 10.1016/j.jbi.2014.01.007] [Medline: 24486355]
77. Castronovo DA, Chui KK, Naumova EN. Dynamic maps: a visual-analytic methodology for exploring spatio-temporal disease patterns. Environ Health 2009 Dec 30;8:61 [FREE Full text] [doi: 10.1186/1476-069X-8-61] [Medline: 20042115]
78. Luo W. Visual analytics of geo-social interaction patterns for epidemic control. Int J Health Geogr 2016 Aug 10;15(1):28 [FREE Full text] [doi: 10.1186/s12942-016-0059-3] [Medline: 27510908]
79. Maciejewski R, Livengood P, Rudolph S, Collins TF, Ebert DS, Brigantic RT, et al. A pandemic influenza modeling and visualization tool. J Vis Lang Comput 2011 Aug;22(4):268-278 [FREE Full text] [doi: 10.1016/j.jvlc.2011.04.002] [Medline: 32288454]
80. Baytas IM, Lin K, Wang F, Jain AK, Zhou J. PhenoTree: interactive visual analytics for hierarchical phenotyping from large-scale electronic health records. IEEE Trans Multimedia 2016 Nov;18(11):2257-2270. [doi: 10.1109/tmm.2016.2614225]
81. Ji X, Chun SA, Geller J. Monitoring Public Health Concerns Using Twitter Sentiment Classifications. In: IEEE International Conference on Healthcare Informatics. 2013 Presented at: CHI'13; Spetember 9-11, 2013; Philadelphia, PA, USA. [doi: 10.1109/ichi.2013.47]
82. Yu Z, Pepe K, Rust G, Ramirez-Marquez J, Zhang S, Bonnet B. Patient-provider Geographic Map: an Interactive Visualization Tool of Patients' Selection of Health Care Providers. In: IEEE Workshop on Visual Analytics in Healthcare. 2017 Presented at: VAHC'17; October 1-11, 2017; Phoenix, AZ, USA. [doi: 10.1109/vahc.2017.8387494]
83. Jinpon P, Jaroensutasinee M, Jaroensutasinee K. Integrated information visualization to support decision-making in order to strengthen communities: Design and usability evaluation. Inform Health Soc Care 2017 Dec;42(4):335-348. [doi: 10.1080/17538157.2016.1255626] [Medline: 28005441]
84. Shaban-Nejad A, Lavigne M, Okhmatovskaia A, Buckeridge DL. PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. Ann N Y Acad Sci 2017 Jan;1387(1):44-53. [doi: 10.1111/nyas.13271] [Medline: 27750378]
85. Afzal S, Maciejewski R, Ebert D. Visual analytics decision support environment for epidemic modeling and response evaluation. In: 2011 IEEE Conference on Visual Analytics Science and Technology. 2011 Presented at: VAST'11; October 23-28, 2011; Providence, RI, USA URL: https://ieeexplore.ieee.org/document/6102457 [doi: 10.1109/vast.2011.6102457]
86. de Mendonça PG, Maciel C, Viterbo J. Visualizing Aedes aegypti infestation in urban areas: A case study on open government data mashups. Int Pol 2015 Aug 31;20(2,3):119-134. [doi: 10.3233/ip-150361]
87. Tate AR, Beloff N, Al-Radwan B, Wickson J, Puri S, Williams T, et al. Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface. J Am Med Inform Assoc 2014;21(2):292-298 [FREE Full text] [doi: 10.1136/amiajnl-2013-001847] [Medline: 24272162]
88. Tilahun B, Kauppinen T, Keßler C, Fritz F. Design and development of a linked open data-based health information representation and visualization system: potentials and preliminary evaluation. JMIR Med Inform 2014 Oct 25;2(2):e31 [FREE Full text] [doi: 10.2196/medinform.3531] [Medline: 25601195]
89. Benis A, Hoshen M. DisEpi: compact visualization as a tool for applied epidemiological research. Stud Health Technol Inform 2017;244:38-42. [Medline: 29039373]

90. Hund M, Böhm D, Sturm W, Sedlmair M, Schreck T, Ullrich T, et al. Visual analytics for concept exploration in subspaces of patient groups: making sense of complex datasets with the doctor-in-the-loop. Brain Inform 2016 Dec;3(4):233-247 [FREE Full text] [doi: 10.1007/s40708-016-0043-5] [Medline: 27747817]
91. Xing Z, Pei J. Exploring disease association from the NHANES data: data mining, pattern summarization, and visual analytics. Int J Data Warehous Min 2010;6(3):11-27. [doi: 10.4018/jdwm.2010070102]
92. Alonso WJ, McCormick BJ. EPIPOI: a user-friendly analytical tool for the extraction and visualization of temporal parameters from epidemiological time series. BMC Public Health 2012 Nov 15;12:982 [FREE Full text] [doi: 10.1186/1471-2458-12-982] [Medline: 23153033]
93. Antoniou D, Georgitsi M, Gkantouna V, Patrinos G, Poulas K, Tsakalidis A, et al. Dautobase: Mining Gems on Autoimmune Diseases Utilizing Web Visualization Technologies. In: Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine. 2010 Presented at: ITAB'10; January 17, 2011; Corfu, Greece URL: https://ieeexplore.ieee.org/document/5687692/authors#authors [doi: 10.1109/itab.2010.5687692]
94. Byrd K, Mansurov A, Baysal O, editors. Mining Twitter Data for Influenza Detection and Surveillance. In: Proceedings of the International Workshop on Software Engineering in Healthcare Systems. 2016 Presented at: SEHS'16; May 2-8, 2016; Austin Texas URL: https://dl.acm.org/doi/abs/10.1145/2897683.2897693 [doi: 10.1145/2897683.2897693]
95. Chen C, Teng Y, Lin B, Fan I, Chan T. Online platform for applying space-time scan statistics for prospectively detecting emerging hot spots of dengue fever. Int J Health Geogr 2016 Nov 25;15(1):43 [FREE Full text] [doi: 10.1186/s12942-016-0072-6] [Medline: 27884135]
96. Chorianopoulos K, Talvis K. Flutrack.org: open-source and linked data for epidemiology. Health Informatics J 2016 Dec;22(4):962-974 [FREE Full text] [doi: 10.1177/1460458215599822] [Medline: 26351261]
97. Garcia-Martí I, Zurita-Milla R, Swart A, van den Wijngaard KC, van Vliet AJ, Bennema S, et al. Identifying environmental and human factors associated with tick bites using volunteered reports and frequent pattern mining. Trans in GIS 2016 May 13;21(2):277-299. [doi: 10.1111/tgis.12211]
98. Frtuni GM, Puflovi D, Stevanoska E, Jevtovi ST, Velinov G, Stoimenov L. Interactive Map Visualization System Based on Integrated Semi-structured and Structured Healthcare Data. In: Data Integration in the Life Sciences. Cham, Switzerland: Springer International Publishing AG; Oct 27, 2017:94-108.
99. Haque W, Urquhart B, Berg E, Dhanoa R. Using business intelligence to analyze and share health system infrastructure data in a rural health authority. JMIR Med Inform 2014 Aug 6;2(2):e16 [FREE Full text] [doi: 10.2196/medinform.3590] [Medline: 25599727]
100. Hardisty F, Klippel A. Analysing spatio-temporal autocorrelation with LISTA-Viz. Int J Geograp Inf Sci 2010 Oct 11;24(10):1515-1526. [doi: 10.1080/13658816.2010.511717]
101. Huang C, Lu R, Iqbal U, Lin S, Nguyen PA, Yang H, et al. A richly interactive exploratory data analysis and visualization tool using electronic medical records. BMC Med Inform Decis Mak 2015 Nov 12;15:92 [FREE Full text] [doi: 10.1186/s12911-015-0218-7] [Medline: 26563282]
102. Ji X, Chun S, Geller J. Epidemic Outbreak and Spread Detection System Based on Twitter Data. In: Health Information Science. Heidelberg, Germany: Springer-Verlag Berlin Heidelberg; 2012:152-163.
103. Shenhui J, Shiaofen F, Bloomquist S, Keiper J, Palakal M, Yuni X. Healthcare Data Visualization: Geospatial and Temporal Integration. In: 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: IVAPP. 2016 Presented at: CCV'16; February 27-29, 2016; Rome, Italy p. 212-219 URL: https://www.scitepress.org/Link.aspx?doi=10.5220/0005714002120219 [doi: 10.5220/0005714002120219]
104. Kaieski N, Oliveira L, Villamil M. Vis-Health: Exploratory Analysis and Visualization of Dengue Cases in Brazil. In: IEEE. 2016 Presented at: 2016 49th Hawaii International Conference on System Sciences (HICSS); 5-8 Jan. 2016; Koloa, HI, USA p. 3063-3072 URL: https://ieeexplore.ieee.org/document/7427568 [doi: 10.1109/hicss.2016.385]
105. Katsis Y, Balac N, Chapman D, Kapoor M, Block J, Griswold W. Big Data Techniques for Public Health: A Case Study. In: International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 2017 Presented at: CHASE'17; July 17-19, 2017; Philadelphia, PA, USA URL: https://ieeexplore.ieee.org/document/8010636 [doi: 10.1109/chase.2017.81]
106. Kruzikas D, Higashi M, Edgar M, Macal C, Graziano D, North M. Using Agent-Based Modeling to Inform Regional Health Care System Investment and Planning. In: 2014 International Conference on Computational Science and Computational Intelligence. 2014 Presented at: CSCI'14; March 10-13, 2014; Las Vegas, NV, USA URL: https://ieeexplore.ieee.org/document/6822331 [doi: 10.1109/csci.2014.120]
107. Maciejewski R, Drake T, Rudolph S, Malik A, Ebert D, editors. Data Aggregation and Analysis for Cancer Statistics - A Visual Analytics Approach. In: 43rd Hawaii International Conference on System Sciences. 2010 Presented at: CSS'10; January 5-8, 2010; Honolulu, HI, USA URL: https://ieeexplore.ieee.org/document/5428311 [doi: 10.1109/HICSS.2010.128]
108. Marek L, Tuček P, Pászto V. Using geovisual analytics in Google Earth to understand disease distribution: a case study of campylobacteriosis in the Czech Republic (2008-2012). Int J Health Geogr 2015 Jan 28;14:7 [FREE Full text] [doi: 10.1186/1476-072X-14-7] [Medline: 25628063]
109. Mitrpanont J, Roungsuriyaviboon J, Sathapornwatanakul T, Sawangphol W, Kobayashi D, Haga J, editors. Extending MedThaiVis-Thai medical research visualization to SAGE2 display walls. In: 22nd International Conference on Information

Technology. 2017 Presented at: INCIT'17; November 2-3, 2017; Nakhonpathom, Thailand URL: https://ieeexplore.ieee.org/abstract/document/8257872 [doi: 10.1109/incit.2017.8257872]

110. Mittelstadt S, Hao M, Dayal U, Hsu M, Terdiman J, Keim D. Advanced Visual Analytics Interfaces for Adverse Drug Event Detection. In: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces. 2014 Presented at: ACI'14; May 11-13, 2014; Como, Italy URL: https://dl.acm.org/doi/proceedings/10.1145/2598153 [doi: 10.1145/2598153.2598156]
111. Ozkaynak M, Dziadkowiec O, Mistry R, Callahan T, He Z, Deakyne S, et al. Characterizing workflow for pediatric asthma patients in emergency departments using electronic health records. J Biomed Inform 2015 Oct;57:386-398 [FREE Full text] [doi: 10.1016/j.jbi.2015.08.018] [Medline: 26327135]
112. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach. Comput Human Behav 2018 Jan;78:98-112 [FREE Full text] [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
113. Perer A, Wang F, Hu J. Mining and exploring care pathways from electronic medical records with visual analytics. J Biomed Inform 2015 Aug;56:369-378 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.020] [Medline: 26146159]
114. Proulx P, Tandon S, Bodnar A, Schroh D, Harper R, Wright W. Avian Flu Case Study with nSpace and GeoTime. In: IEEE Symposium On Visual Analytics Science And Technology. 2006 Presented at: IEEE'06; October 31-November 2, 2006; Baltimore, MD, USA URL: https://ieeexplore.ieee.org/document/4035744 [doi: 10.1109/vast.2006.261427]
115. Soulakis ND, Carson MB, Lee YJ, Schneider DH, Skeehan CT, Scholtens DM. Visualizing collaborative electronic health record usage for hospitalized patients with heart failure. J Am Med Inform Assoc 2015 Mar;22(2):299-311 [FREE Full text] [doi: 10.1093/jamia/ocu017] [Medline: 25710558]
116. Toddenroth D, Ganslandt T, Castellanos I, Prokosch H, Bürkle T. Employing heat maps to mine associations in structured routine care data. Artif Intell Med 2014 Feb;60(2):79-88. [doi: 10.1016/j.artmed.2013.12.003] [Medline: 24389331]
117. Torres S, Eicher-Miller H, Boushey C, Ebert D, Maciejewski R. Applied Visual Analytics for Exploring the National Health and Nutrition Examination Survey. In: 45th Hawaii International Conference on System Sciences. 2012 Presented at: CSS'12; January 4-7, 2012; Maui, HI, USA URL: https://ieeexplore.ieee.org/document/6149111 [doi: 10.1109/hicss.2012.116]
118. Yan W, Palm L, Lu X, Nie S, Xu B, Zhao Q, et al. ISS--an electronic syndromic surveillance system for infectious disease in rural China. PLoS One 2013;8(4):e62749 [FREE Full text] [doi: 10.1371/journal.pone.0062749] [Medline: 23626853]
119. Jankun-Kelly TJ, Ma KL, Gertz M. A model and framework for visualization exploration. IEEE Trans Vis Comput Graph 2007;13(2):357-369. [doi: 10.1109/TVCG.2007.28] [Medline: 17218751]
120. Fayyad U, Piatetsky-Shapiro G, Smyth P. The KDD process for extracting useful knowledge from volumes of data. Commun ACM 1996 Nov;39(11):27-34. [doi: 10.1145/240455.240464]
121. Tu K, Mitiku TF, Ivers NM, Guo H, Lu H, Jaakkimainen L, et al. Evaluation of electronic medical record administrative data linked database (EMRALD). Am J Manag Care 2014;20(1):e15-e21. [Medline: 24669409]
122. Gao S, Rao J, Kang Y, Liang Y, Kruse J. Mapping county-level mobility pattern changes in the United States in response to COVID-19. SIGSPATIAL Special 2020 Jun 3;12(1):16-26. [doi: 10.1145/3404111.3404115]
123. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
124. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the 'statistical analyses and methods in the published literature' or the SAMPL guidelines. Int J Nurs Stud 2015 Jan;52(1):5-9. [doi: 10.1016/j.ijnurstu.2014.09.006] [Medline: 25441757]
125. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
126. Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, et al. Checklist for Analytical Cross Sectional Studies. The Joanna Briggs Institute. 2017. URL: https://joannabriggs.org/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Analytical_Cross_Sectional_Studies2017_0.pdf [accessed 2020-11-14]
127. Howard Z, Somerville MM. A comparative study of two design charrettes: implications for codesign and participatory action research. CoDesign 2014 Mar 6;10(1):46-62. [doi: 10.1080/15710882.2014.881883]
128. Social Determinants of Health. World Health Organization. 2020. URL: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1 [accessed 2020-11-14]
129. McGinnis JM, Stuckhardt L, Saunders R, Smith M. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington DC, USA: National Academies Press; 2013.
130. Etheredge LM. A rapid-learning health system. Health Aff (Millwood) 2007;26(2):w107-w118. [doi: 10.1377/hlthaff.26.2.w107] [Medline: 17259191]
131. Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. Ann Intern Med 2012 Aug 7;157(3):207-210. [doi: 10.7326/0003-4819-157-3-201208070-00012] [Medline: 22868839]
132. Basole RC, Braunstein ML, Sun J. Data and analytics challenges for a learning healthcare system. J Data Inform Qual 2015 Jul 21;6(2-3):1-4. [doi: 10.1145/2755489]

133. Berry I, Soucy JR, Tuite A, Fisman D, COVID-19 Canada Open Data Working Group. Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. Can Med Assoc J 2020 Apr 14;192(15):E420 [FREE Full text] [doi: 10.1503/cmaj.75262] [Medline: 32392510]
134. McGowan J, Sampson M, Salzwedel D, Cogo E, Foerster V, Lefebvre C. Press peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
135. Keshavjee K, Mirza K, Martin K. The next generation EMR. Stud Health Technol Inform 2015;208:210-214. [Medline: 25676975]
136. Mühlbacher T, Piringer H. A partition-based framework for building and validating regression models. IEEE Trans Vis Comput Graph 2013 Dec;19(12):1962-1971. [doi: 10.1109/TVCG.2013.125] [Medline: 24051762]
137. Potter K, Wilson A. Ensemble-Vis: A Framework for the Statistical Visualization of Ensemble Data. In: IEEE International Conference on Data Mining Workshops. 2009 Presented at: ICDMW'09; December 6, 2009; Miami, FL, USA. [doi: 10.1109/ICDMW.2009.55]
138. Sedlmair M, Heinzl C, Bruckner S, Piringer H, Möller T. Visual parameter space analysis: a conceptual framework. IEEE Trans Vis Comput Graph 2014 Dec;20(12):2161-2170. [doi: 10.1109/TVCG.2014.2346321] [Medline: 26356930]
139. Harrower M. Tips for designing effective animated maps. CP 2003 Mar 1(44):63-65. [doi: 10.14714/CP44.516]
140. Shahar Y, Musen MA. Knowledge-based temporal abstraction in clinical domains. Artif Intell Med 1996 Jul;8(3):267-298. [doi: 10.1016/0933-3657(95)00036-4] [Medline: 8830925]
141. Müller E, Günnemann S, Assent I, Seidl T. Evaluating clustering in subspace projections of high dimensional data. Proc VLDB Endow 2009 Aug;2(1):1270-1281. [doi: 10.14778/1687627.1687770]
142. Shneiderman B. The Eyes Have It: a Task by Data Type Taxonomy for Information Visualizations. In: IEEE Symposium on Visual Languages. 1996 Presented at: VL'96; September 3-6, 1996; Boulder, CO, USA. [doi: 10.1109/VL.1996.545307]
143. Wasi P. Triangle That Moves the Mountain and Health Systems Reform Movement in Thailand. New York, USA: Health Systems Research Institute; 2000.
144. Bryant T. Social Determinants of Health: The Canadian Facts. New York, USA: Routedge; 2010.
145. Tate AR. Developing Quality Scores for Electronic Health Records for Clinical Research: a Study Using the General Practice Research Database. In: Proceedings of the First International Workshop on Managing Interoperability and Complexity in Health System. 2011 Presented at: WMI'11; April 7-9, 2011; New York, USA.
146. Volz J. Silk-a link discovery framework for the web of data. Ldow 2009;53:538 [FREE Full text]
147. High Performance Visualization. New York, USA: Chapman and Hall; 2012.
148. Tableau. URL: https://www.tableau.com/ [accessed 2020-11-24]
149. Qlik. URL: https://www.qlik.com/us/ [accessed 2020-11-24]
150. Power BI. Microsoft. URL: https://powerbi.microsoft.com/en-us/ [accessed 2020-11-24]

Abbreviations

API: application programming interface
EHR: electronic health record
EMR: electronic medical record
GIS: geographic information system
HSR: health services research
MeSH: Medical Subject Headings
ML: machine learning
NLP: natural language processing
PRESS: Peer Review of Electronic Search Strategies
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
VA: visual analytics

Edited by G Eysenbach; submitted 20.01.20; peer-reviewed by R Dias, L Sheets, R De Croon; comments to author 21.03.20; revised version received 01.07.20; accepted 24.09.20; published 03.12.20.

Please cite as:
Chishtie JA, Marchand JS, Turcotte LA, Bielska IA, Babineau J, Cepoiu-Martin M, Irvine M, Munce S, Abudiab S, Bjelica M, Hossain S, Imran M, Jeji T, Jaglal S
Visual Analytic Tools and Techniques in Population Health and Health Services Research: Scoping Review
J Med Internet Res 2020;22(12):e17892
URL: https://www.jmir.org/2020/12/e17892
doi:10.2196/17892
PMID:33270029

Review

Effectiveness, Acceptability, and Feasibility of Digital Health Interventions for LGBTIQ+ Young People: Systematic Review

Dylan Gilbey[1,2], BSc (Hons); Helen Morgan[1,3], BA (Hons); Ashleigh Lin[1,4], BSc (Hons), MPsych, PhD; Yael Perry[1,4], BPsych (Hons), MPsych, PhD

[1]Telethon Kids Institute, Perth, Australia

[2]School of Psychological Science, The University of Western Australia, Perth, Australia

[3]Discipline of Psychology, College of Science, Health, Engineering and Education, Murdoch University, Perth, Australia

[4]Centre for Child Health Research, The University of Western Australia, Perth, Australia

Corresponding Author:
Dylan Gilbey, BSc (Hons)
Telethon Kids Institute
15 Hospital Avenue
Perth
Australia
Phone: 61 63191195
Email: Dylan.Gilbey@telethonkids.org.au

Abstract

Background: Young people (aged 12-25 years) with diverse sexuality, gender, or bodily characteristics, such as those who identify as lesbian, gay, bisexual, transgender, intersex, or queer (LGBTIQ+), are at substantially greater risk of a range of mental, physical, and sexual health difficulties compared with their peers. Digital health interventions have been identified as a potential way to reduce these health disparities.

Objective: This review aims to summarize the characteristics of existing evidence-based digital health interventions for LGBTIQ+ young people and to describe the evidence for their effectiveness, acceptability, and feasibility.

Methods: A systematic literature search was conducted using internet databases and gray literature sources, and the results were screened for inclusion. The included studies were synthesized qualitatively.

Results: The search identified 38 studies of 24 unique interventions seeking to address mental, physical, or sexual health–related concerns in LGBTIQ+ young people. Substantially more evidence-based interventions existed for gay and bisexual men than for any other population group, and there were more interventions related to risk reduction of sexually transmitted infections than to any other health concern. There was some evidence for the effectiveness, feasibility, and acceptability of these interventions overall; however, the quality of evidence is often lacking.

Conclusions: There is sufficient evidence to suggest that targeted digital health interventions are an important focus for future research aimed at addressing health difficulties in LGBTIQ+ young people. Additional digital health interventions are needed for a wider range of health difficulties, particularly in terms of mental and physical health concerns, as well as more targeted interventions for same gender–attracted women, trans and gender-diverse people, and people with intersex variations.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020128164; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=128164

(J Med Internet Res 2020;22(12):e20158) doi:10.2196/20158

KEYWORDS

systematic review; mental health; physical health; sexual health; youth; sexuality; gender; mobile phones

Introduction

Young people who are lesbian, gay, bisexual, transgender, intersex, queer and other people of diverse sexuality, gender, or bodily characteristics (LGBTIQ+) are known to experience a range of disparities in health outcomes compared with their peers [1]. These include higher rates of mental health difficulties, such as depression and suicidality [2-4]; physical and sexual health problems, such as the incidence of HIV [5]; cigarette and alcohol use [6-8], obesity [9], and teen pregnancy [10]. Later in adulthood, the confluence of these health issues conveys further risk for cancer [11] and cardiovascular disease diagnoses [12-14]. The burden of disease that these disparities carry is a public health issue that urgently requires safe, effective, and early intervention.

These health disparities are compounded by barriers that negatively impact the ability of LGBTIQ+ young people to access health services that are safe and adequately meet their needs. Young people in the general population face many barriers to help seeking, including inadequate resources and lack of accessibility, desire for self-reliance, and anticipated stigma for reporting certain health difficulties such as mental illnesses or HIV [15,16]. LGBTIQ+ young people face a range of additional difficulties, such as low parental support, which can lead to homelessness [17-19], and unemployment due to discrimination [20,21], which may intensify these help-seeking barriers [22]. This group also faces unique help-seeking difficulties, such as anticipated and experienced stigma surrounding their identities [22-24], concerns about disclosure or their compromised confidentiality [24], and low perceived confidence in the ability of service providers to deliver LGBTIQ+ supportive care [23-26]. These problems may be particularly pronounced among queer and trans and gender-diverse young people [25,27]. The factors impeding health care access and use may significantly contribute to the maintenance of health disparities in LGBTIQ+ young people [28,29]. Accordingly, removing these obstacles is an important step toward improving health in this population.

Digital health interventions, such as those delivered via computers, websites, smartphones, or tablets, have been identified as an important potential avenue to improve health care access and use among young people in this group [30-33]. Accessing support digitally allows young people to bypass many of the aforementioned barriers. Anonymity facilitates private access to support and minimizes stigma [33]. In addition, digital health interventions confer further benefits beyond traditional clinical environments, being available without travel, accessible at all hours, and having no waitlists [34]. Self-guided digital health interventions are also cost- and resource-effective to access and disseminate, requiring less direct input from clinicians [34], giving consumers greater control and empowerment regarding their own health needs [35]. These considerations are especially pertinent for widening the support available to populations that are restricted from accessing traditional health care services [36]. However, digital health interventions are also commonly reported to have very high rates of attrition and disengagement [37,38], with up to 60% to 80% of users discontinuing their use [39-41], and the quality of evidence supporting the effectiveness of digital health interventions is also frequently lacking [42-45]. Along with the limited availability of many evidence-based digital health interventions beyond the research context, these issues call into question the real-world utility of these interventions despite their proposed theoretical benefits.

LGBTIQ+ young people are adept and frequent users of digital technologies [33,46,47]; however, research indicates that technology use can present several challenges. Evidence suggests that the internet (including social media and online communities in particular) can be harmful in this population (as well as young people more broadly) [31] due to the potential for toxic interactions and exposure to distressing content [48], such as the normalization of self-harm and suicidal behaviors [49]. However, an array of benefits associated with technology use has also been documented in LGBTIQ+ young people. The internet allows LGBTIQ+ young people to explore their identities in an anonymous and safe way, seek out peers who share their identities, and come out to others in a low-risk environment [47]. The internet also facilitates an important component of the sexual development of LGBTIQ+ young people, enabling the exploration of same-gender attraction for some and seeking out romantic or sexual partners [50]. LGBTIQ+ young people may also already use the internet to access resources that are relevant and safe for them [47], suggesting that digital health interventions targeting this group may be useful. The concept of digital delivery of interventions is generally acceptable to this group, and this is particularly true when they are specifically targeted with LGBTIQ+ themes [32]. Many existing digital health interventions are not specifically applicable to LGBTIQ+ people [51]; however, untailored interventions may exacerbate feelings of alienation [31,52,53]. Given these factors, the development and evaluation of targeted digital health interventions for LGBTIQ+ young people may represent an opportunity to improve the delivery of health care to this group, should the benefits outweigh the known challenges associated with digital health care delivery discussed above.

In response, there has been a rapid increase in the number of such interventions over the past decade. To date, however, there has not been a comprehensive review summarizing the scope and use of digital health interventions that currently exist for this population. Knight et al [54] recently reviewed web-based interventions for HIV or sexually transmitted infection (STI) risk reduction in young men who have sex with men (MSM); however, this review did not capture digital health interventions that are delivered through other digital modalities (eg, mobile apps), those that address other health issues, or those that target women, gender-diverse individuals, or individuals with intersex variations. There have also been several recent reviews focused on the mental health of LGBTIQ+ young people [55,56] and adults [57], which have referenced a combined total of 4 digital interventions for this population. However, these reviews were not explicitly focused on the use of digital technology nor did they consider interventions designed to improve a broad range of health outcomes in this population.

A more extensive summary of this rapidly growing field of research will assist in identifying gaps in the development of interventions and determining the overall evidence for their use across the full diversity of the young LGBTIQ+ community. Therefore, this review aims to answer the following questions: (1) What are the characteristics of evidence-based digital health interventions for improving mental, physical, and sexual health outcomes in LGBTIQ+ young people? (2) Are targeted digital health interventions effective at improving health outcomes in LGBTIQ+ young people? (3) Are targeted digital health interventions acceptable and feasible for LGBTIQ+ young people?

Methods

Protocol and Registration

The protocol for this review was registered using PROSPERO (Prospective Register of Systematic Reviews; ID CRD42020128164) in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) recommendations [58].

Eligibility Criteria

Types of Participants

The population of interest was LGBTIQ+ young people. The LGBTIQ+ term was used in its broadest sense to capture young people of diverse sexuality (including but not limited to those that identify as gay, lesbian, bisexual, or pansexual), diverse gender (including but not limited to those who identify as trans or nonbinary), with diverse sex characteristics (including but not limited to people with intersex variations), or people falling across any combination of these categories. The search strategy (below) was designed to be as inclusive as possible of the wide variety of identities that LGBTIQ+ people may hold including, for example, people who fall within these aspects of diversity without explicitly identifying as such (eg, MSM). *Young people* was defined as being primarily people between the ages of 12 and 25 years; the mean age of study participants was required to fall within this range for a study to be eligible to be included in the review. In addition, studies were required to have intentionally and specifically recruited young people.

Types of Intervention

The review focused on interventions designed to effect change through predominantly digital means (eg, using a computer, website, tablet, or smartphone). To be included, interventions were required to be targeted or intended to specifically effect change in health outcomes in LGBTIQ+ people. Interventions delivered via telephone with no technological function or an implantable device that is remotely monitored were excluded. Interventions were also required to have minimal human guidance in the intervention itself if present at all. Specifically, the action, process of intervening, or behavior change techniques must have been delivered by the digital technology itself not a health professional working over a digital medium. This criterion was implemented because, in interventions that blend digital and human support, the impact of the intervention cannot be meaningfully attributed to the digital component alone [59].

The judgment of the level of human guidance was made by considering the ratio of clinicians or staff to users, and the centrality of the human guidance to effecting change in the health outcome, which itself relied on factors such as the ratio of guided versus unguided time during the intervention. No hard limits on these factors were set prior to conducting the review, as making this judgment required the full context of the intervention to be considered holistically. A judgment about the duration of the human guidance, for example, could not meaningfully be made without consideration of the purpose of that period of guidance and how it fits into the goals and process of the intervention as a whole. What was counted as *minimal human guidance* was therefore determined on a case-by-case basis requiring consensus from the reviewers.

Types of Studies

To be included, studies should have conducted an evaluation of a specific intervention as described above. Evaluation in some form was required in keeping with standards of evidence-based practice. The term *evaluation* was inclusive of examination of efficacy, effectiveness, acceptability, or feasibility, with a minimum of any one form of evaluation required for inclusion. All quantitative, qualitative, or mixed methods studies were eligible for inclusion. Comparators or control groups were not necessary for inclusion in the review. Protocols describing an intervention without any evaluation were not included in the review; however, when protocols were identified, steps were taken to determine if the corresponding data were publicly available. Studies evaluating the concept of digital health interventions generally or studies describing the initial development of an intervention were also ineligible. Studies evaluating digital health interventions not specifically designed for LGBTIQ+ young people were also excluded, even if conducted with an LGBTIQ+ sample.

No specification was made for the location of the study; however, studies were required to be published in the English language. With the aim of reducing the risk of publication bias [60], gray literature was considered eligible and studies were not required to be peer reviewed to be included in the review. The search was restricted to studies published after January 1, 2000. This cut-off was selected because of the types of interventions eligible for this review, only web-based interventions may have existed at this time, and the likelihood of such an intervention existing specifically tailored for a select, marginalized group was deemed to be extremely low. Scoping searches conducted before deciding on this cut-off did not determine any evidence of existing interventions contrary to this conclusion.

Types of Outcomes

The review was designed to capture interventions seeking to improve health outcomes or to prevent negative health outcomes. This was inclusive of mental health outcomes (eg, symptoms or diagnoses of mental disorders, well-being, distress), physical health outcomes (eg, smoking, weight loss), or sexual health outcomes (eg, pre-exposure prophylaxis [PrEP] adherence, condom use). Any outcome reasonably perceived to represent some aspect of health and well-being was considered relevant to the review. For studies evaluating efficacy, changes must have been reported in measures of at least one of these outcomes. For studies evaluating acceptability or feasibility, at least one index of these factors (eg, surveys of participant experiences, adherence, or attrition rates) must have been reported.

Search Strategy

Internet databases such as PsycINFO (Ovid) and MEDLINE (Ovid) were systematically searched on August 13, 2019, and potentially relevant peer-reviewed publications were extracted. These searches were conducted using a combination of subject headings and keywords corresponding to the following themes:

LGBTIQ+, Youth/Young People (aged 12-25 years), Mental Health, Physical Health, Sexual Health, Digital and Intervention. The search terms for LGBTIQ+, Youth/Young People, and Mental Health themes were adapted from those previously reported by Gilbey et al [61] and Lee et al [62]. The search terms for the other themes were devised from a broad initial scoping search of relevant articles to identify key terms and phrases. The search strategy for PsycINFO (Ovid) is presented in Multimedia Appendix 1.

Additional searches were made using Scopus, ProQuest Dissertations, Google, Google Scholar, OpenGrey, WorldCat, ClinicalTrials.gov, and JMIR Publications, during July and August 2019. Each of these searches were conducted with several simple keyword searches (eg, *LGBTQ*, *transgender*) as the relatively low number of relevant articles available in each source made a more comprehensive, and therefore restrictive, search process unnecessary. The reference lists of several other related reviews and key articles on the subject were also hand-searched for potentially relevant articles during July 2019. Google Scholar alerts were monitored for any additional articles published until April 2020. The searches of PsycINFO (Ovid) and MEDLINE (Ovid) were repeated on March 25, 2020, and articles published since August 2019 were manually searched for any newly published studies.

Screening

The titles and abstracts of the articles identified by the search were screened for relevance by the lead author (DG), removing articles with no clear relevance to the topic of the review. Two authors (DG and HM) then screened the full-text of the remaining articles independently, with differences in opinion resolved in discussion with a third author (YP) in which full agreement was sought.

Data Extraction

The following data items were extracted from eligible studies: author(s), year, participant age (mean and range), description of sample (eg, LGBTIQ+ status), sample size, study design, study setting, intervention type, content and delivery, digital platform, control group type (if relevant), degree of human guidance in the intervention, health outcome(s), acceptability outcome(s), and feasibility outcome(s). A second reviewer (HM) cross-checked these data.

Critical Appraisal

Following data extraction, studies were evaluated using the Mixed Methods Appraisal Tool (MMAT) [63] by 2 reviewers (DG and HM). The MMAT is designed for the assessment of methodological quality of studies with a range of designs (qualitative, quantitative, and mixed methods), such as those reported herein. On the basis of the recommendations of the authors of the tool, quantitative quality scores were not derived; instead, the results of the appraisal are discussed qualitatively.

Synthesis of Results

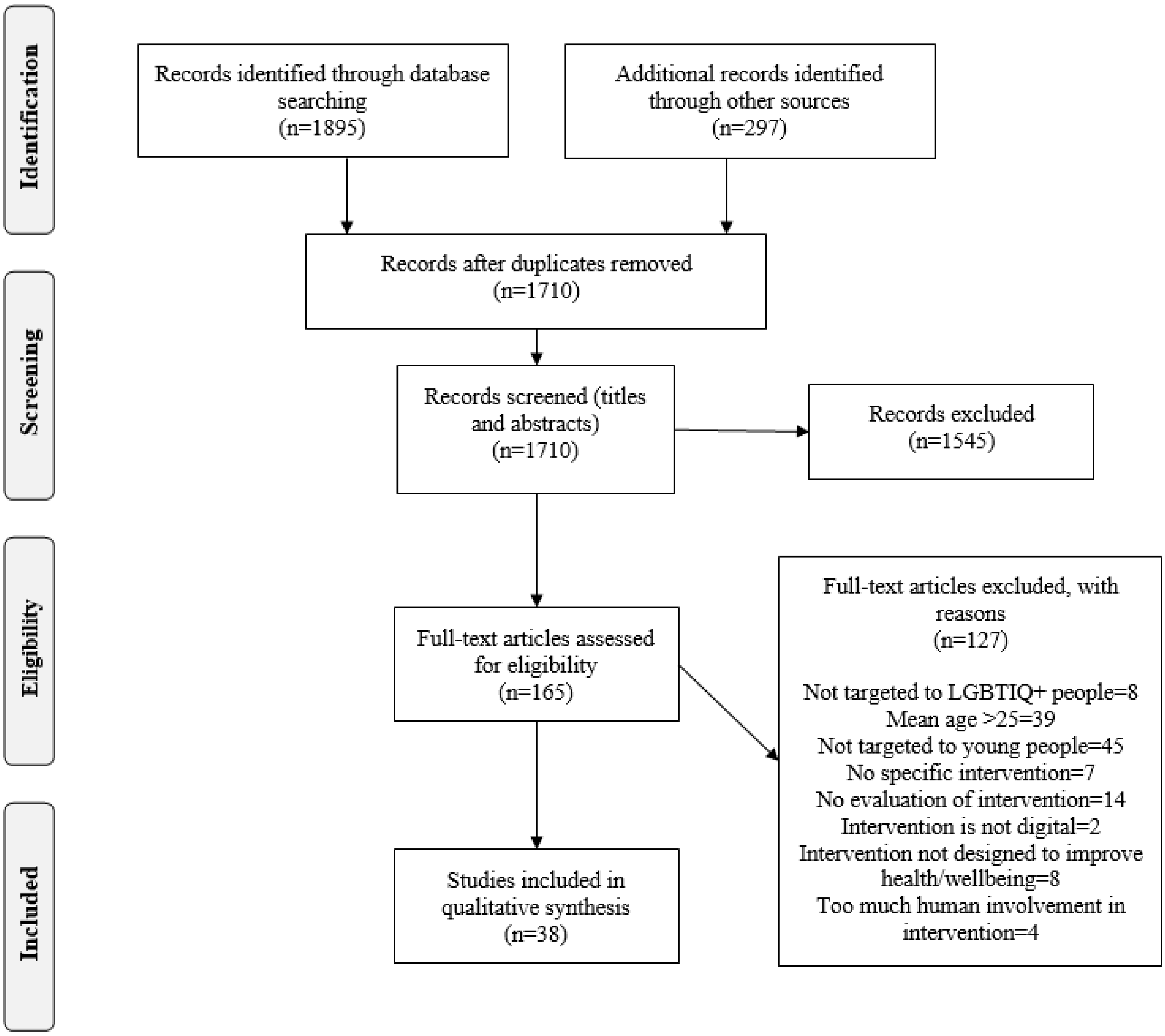
Owing to the wide array of interventions, targets of intervention, intervention modality, and health outcomes measured, it was anticipated that a quantitative synthesis (of those studies reporting quantitative data) would be neither feasible nor informative. Therefore, the results of the studies were synthesized qualitatively.

Results

Study Selection

The search and screening process is displayed in a PRISMA flow chart in Figure 1. A total of 2192 studies were identified in the search. Following title and abstract screening, 165 studies were retained for full-text screening. The final number of studies retained for the review following full-text screening was 38.

Figure 1. Study flowchart. LGBTIQ+: lesbian, gay, bisexual, transgender, intersex, queer.



Study Characteristics

The 38 studies captured the results of studies examining 24 unique interventions conducted across 3 countries (the United States, the United Kingdom, and New Zealand). Of these, 5 targeted primarily mental health–related issues, one targeted primarily physical health–related concerns, one targeted primarily sexual health and well-being, and 17 targeted risk reduction or management of STIs. A total of 22 interventions focused specifically on young people who are attracted to the same gender (referred to with a variety of terms, eg, sexual minority, lesbian/gay/bisexual/queer people, MSM), of which 19 interventions were described as being focused on young men (eg, gay/bisexual men, MSM). Several studies that described their target audience as MSM also included trans women under this descriptor. One study targeted people who identify as lesbian, gay, bisexual, transgender and queer (LGBTQ) generally [64] and one intervention targeted transgender men and nonbinary people specifically [65]. A total of 3 interventions specifically targeted young LGBTQ+ people of color, although several other interventions were also conducted with predominantly non-White participants. No studies were identified that sought to improve health in young people with intersex variations, and no studies were identified that sought to improve health in young women alone. The characteristics of the included studies are reported in Table 1. Because of the length of the table, the characteristics of the included studies addressing sexually transmitted infection risk reduction and management are reported separately in Multimedia Appendix 2 [66-91]. Brief summaries of each intervention as well as their core findings are also provided in Multimedia Appendix 3 [64-100].

Table 1. Summary of digital mental, physical and sexual health interventions for lesbian, gay, bisexual, transgender, intersex or queer young people.

Intervention category and name	Primary health outcome	Study	Participant mean age (range; intervention condition)	Description of sample (eg, LGBTIQ[a] status/identification)	Sample size	Study design	Study setting	Intervention type	Digital platform	Primary measured constructs and effects
Drug abuse prevention										
Unnamed intervention	Drug abuse	Schwinn et al [92]	16.1 (15-16)	Sexual minority youth. Same-sex attracted=90, both-sex attracted=116, opposite-sex attracted=14, not sure= 13	236	Quantitative, RCT[b] and follow-up, efficacy	United States	Interactive skill-building sessions	Computer via web	• Alcohol use _[c] • Marijuana use – • Cigarette use – • Peer drug use ↓[d] • Other drug use ↓
Smoking cessation										
Put It Out Project	Smoking cessation	Vogel et al [98]	19.7 (18-25)	Sexual and gender minority young adults. Gay/lesbian=6, Bisexual=15, Queer=2, Pansexual=8, nonbinary=10, Trans=2	27	Mixed methods, acceptability and feasibility	United States	Social media (Facebook)	Web	N/A[e]
	Smoking cessation	Vogel et al [99]	21.4 (18-25)	Sexual and gender minority young adults. Gay=29, lesbian=30, bisexual/pansexual=93, other=13	165	Quantitative, RCT and follow-up, efficacy, and acceptability	N/A	Social media (Facebook)	Web	• Number of cigarettes smoked weekly ↓ • Self-reported smoking abstinence ↑[f] • Biochemically verified smoking abstinence ↑
Internalizing disorder prevention/management										
Rainbow SPARX	Internalizing symptoms (depression and anxiety)	Lucassen [93]; thesis, studies two and three described in published articles below	N/A[e]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Internalizing symptoms (depression and anxiety)	Lucassen et al [94]	16.5 (13-19)	Sexual minority youth	21	Quantitative, uncontrolled pilot, acceptability and feasibility testing	New Zealand	Serious game	Computer via CD	• Depressive symptoms ↓ • Anxiety symptoms ↓

Intervention category and name	Primary health outcome	Study	Participant mean age (range; intervention condition)	Description of sample (eg, LGBTIQ[a] status/identification)	Sample size	Study design	Study setting	Intervention type	Digital platform	Primary measured constructs and effects
	Internalizing symptoms (depression and anxiety)	Lucassen et al [94]	16.4 (13-19)	Sexual minority youth	25	Qualitative, acceptability testing	New Zealand	Serious game	Computer via CD	N/A
	Internalizing symptoms (depression and anxiety)	Lucassen et al [53]	17.9 (15-22)	LGBT+ youth and health professionals	21 youth and 6 professionals	Qualitative, acceptability testing	United Kingdom	Serious game	Computer via CD	N/A
TODAY!	Internalizing symptoms (depression and anxiety)	Fleming et al [96]	19.0 (18-20)	Young sexual minority men. Gay=9	9	Qualitative, usability testing	United States	Mobile app	Mobile phone	N/A
Nonspecific mental health interventions										
Unnamed intervention	Psychological distress	Pachankis and Goldfried [97]	20.2 (Range not provided)	Gay male college students	77	Quantitative, RCT and follow-up, efficacy	United States	Expressive writing	PC	- Depressive symptoms – - Psychological well-being –
QueerViBE	Psychological well-being	Martin [65]	18.0 (15-21)	Young trans men and nonbinary people. Trans male=89, nonbinary=50, questioning=5, other=12	156	Mixed methods, RCT, interviews, acceptability and efficacy	United Kingdom	YouTube videos	Web	- Psychological distress ↓
Sexual health and wellbeing										
Queer Sex Ed	Sexual health	Mustanski et al [64]	17.9 (16-20)	LGBT young people. Gay/lesbian=142, bisexual=31, queer=27, unsure/questioning=2, transgender=14	202	Mixed methods, one-arm pilot, acceptability, feasibility and efficacy	United States	Web-based curriculum	Computer via web	- Sexual functioning ↑ - HIV knowledge ↑ - STD knowledge[g] ↑ - Contraceptive methods knowledge ↑ - HIV testing location awareness ↑

---

[a] LGBTIQ: lesbian, gay, bisexual, transgender, intersex, or queer.

[b] RCT: randomized controlled trial.

[c] –: No change.

[d] ↓: Significant decrease.

[e] N/A: not applicable.

[f] ↑: Significant increase.

[g] STD: sexually transmitted disease.

Risk of Bias in Individual Studies

Mixed Methods Studies

Mixed methods designs were used well among the included studies overall, with most meeting all of the criteria in the MMAT for such designs. This is not to say that these studies did not have methodological flaws, as the MMAT mixed methods subsection does not generally consider the quality of the individual qualitative and quantitative components but rather their intersection and integration. These components of the studies were also, therefore, considered individually and are included among those described below.

Quantitative Studies

The majority of included quantitative trial studies were described as pilot or feasibility studies (17 out of 27), and their methodological quality was lacking in many cases. In total, 15 of the 27 trials were randomized trials, of which 4 did not report appropriate randomization procedures, 12 did not report blinding procedures, and participant adherence to the intervention was only reported in 2 studies. Of the 12 nonrandomized trials, 10 did not report accounting for confounding variables in their design and analysis; however, representativeness in the study samples was generally adequate. The majority of the 12 quantitative descriptive studies' methodologies were vulnerable to nonresponse bias.

Qualitative Studies

Studies' methodologies were generally sufficient to meet the MMAT criteria for qualitative studies. It is worth noting, however, that most of the studies did not specifically outline the methodological framework underpinning the study, and it is unclear whether this was due to inadequate reporting or the absence of such structures entirely. Some studies did not appear to adequately substantiate their conclusions with the data, but again it was difficult to interpret whether this was due to flaws in methodology or omission of reporting.

Synthesis of Results

Content of Digital Health Interventions for LGBTIQ+ Young People

Of the 5 interventions targeting mental health difficulties, 2 focused on internalizing symptom reduction [93-95], 1 targeted drug abuse prevention [92], and 2 focused on nonspecific aspects of psychological well-being [65,97]. Although few in number, existing digital mental health interventions targeted several LGBTIQ+ subgroups, and there was relatively little conceptual overlap between them. Four of these interventions were theory-driven, and only one mental health intervention showed noteworthy community involvement in the development of the intervention. With one exception, mental health interventions tended to rely on some form of skill-building or otherwise didactic content delivery.

There were few interventions targeting physical health problems or sexual health and well-being in LGBTIQ+ young people. Only one digital intervention focused on smoking cessation and targeted physical health in LGBTIQ+ young people [98,99]. This intervention showed aspects of both community-driven and theoretical designs. One digital intervention targeting sexual and reproductive well-being overall in LGBTIQ+ young people was developed with a theoretical framework but no community input [64]. Some aspects of this intervention overlapped with interventions targeting the risk of STIs, such as increasing condom use; however, other aspects diverged, such as including content on healthy relationships more broadly.

The majority of the interventions identified in this review were targeted toward risk reduction or the management of STIs. Of the 17 interventions identified as having focused on risk reduction and management of STIs, 1 focused on pre-exposure prophylaxis adherence [66], 2 focused on reducing unprotected sex [67-71], 7 targeted multiple aspects of HIV prevention [72-83], 2 focused on HPV prevention [84,85,101], 3 focused on increasing STI testing [86-89], and 2 focused on antiretroviral medication adherence [90,91]. All 17 interventions were targeted toward young men who are attracted to men, 2 of which were specifically for adolescents. The majority of these interventions included some aspect of community involvement in their design, although the extent varied from iterations based on user feedback to more central participatory design, which few interventions involved. All but one intervention was theory-driven in nature, with most interventions drawing from either the Information-Motivation-Behavioral Skills Model or Social Cognitive Theory in their design. Therefore, there was notable conceptual overlap among these interventions, which largely varied only in their delivery format and the breadth of their focus (ie, targeting specific aspects of risk reduction, such as condom use, versus a variety of such behaviors).

Delivery of Digital Health Interventions for LGBTIQ+ Young People

The most common platforms for digital health interventions were websites and mobile apps that, combined, represented over half of the interventions identified. A smaller number of interventions were delivered via computer software. Many interventions used gamification, or elements of game playing such as point scoring, in their delivery; however, only 2 interventions were fully gamified in nature [90,95]. Few interventions incorporated social interactivity and where present, they were typically minimal [72,75,78,90,98]. Most interventions were multimedia, incorporating a number of different delivery formats and types of content.

The vast majority of interventions delivered information to effect change either in terms of building awareness about the health issue in question or teaching skills to enable behavior change. This information is typically delivered via text or videos. For some interventions, this was the entirety of their scope; however, others included quizzes, games, or practice scenarios to consolidate the knowledge being presented. Interventions varied significantly in the extent of their personalization; some interventions delivered the same content to all participants, while others provided opportunities for personalized input and then delivered information specific to the individual's situation or needs at the time.

Interventions varied in the duration and intensity of their delivery. In total, 16 of the 24 interventions appeared to be intended to be a perpetually available resource that could be accessed at any time and largely completed in a single instance if desired. Intervention duration ranged from very brief completion times, as low as 10 min [84], to up to 3.5 hours [95] to complete. In total, 8 of the 24 interventions staggered their delivery in some respect, such as presenting new content over a certain period [66,90], or otherwise incentivizing users to return to the intervention over a period of up to several months [96]. Few interventions described periods of use over 2 months; however, this may reflect the limitations of trial periods rather than their ideal dissemination. Overall, a small minority of interventions appear to be available for public use at present.

Effectiveness of Digital Health Interventions for LGBTIQ+ Young People

With regard to the effectiveness of these interventions, there was consistent evidence from a number of interventions that digital health interventions could improve HIV testing rates in young MSM [79,80,88,89]. Aside from this, STI-focused interventions appeared to effect change more consistently in cognitive or attitudinal outcomes, such as HIV awareness and preparedness to use condoms, than in behavioral outcomes in practice, such as condom use and unprotected anal sex. Some studies observed changes in these outcomes [69,73,74,83], while others did not [67,79,80]. There was insufficient evidence to suggest that interventions targeting a specific risk-related outcome were more effective at improving that outcome than interventions that sought change in a variety of risk-related outcomes.

Comparatively, few interventions were targeted toward improvement of mental or physical health issues; however, the interventions that did exist in this sphere were more targeted and overlapped less in scope. With the exception of an expressive writing intervention [97], digital health interventions demonstrated preliminary effectiveness in reducing internalizing symptoms such as depression and psychological distress [65,95]. Digital health interventions have also been reported to be effective at reducing substance use, including cigarette use [99] and peer drug use [92]. Owing to the limited number of interventions targeting these difficulties, it is difficult to determine any patterns regarding the factors predicting greater effectiveness; however, the only intervention that did not show notable effectiveness was also the only intervention to not present any didactic or skill-building content to the user.

Given that the majority of the interventions included in the review were multimedia in some form, it was not possible to draw conclusions about the delivery aspects that would most reliably effect change in outcomes. Furthermore, and importantly, given that the overall quality of the interventions included in this review was suboptimal, their effectiveness must be viewed in light of limitations associated with methodology and reporting.

Acceptability and Feasibility of Digital Health Interventions for LGBTIQ+ Young People

Overall, digital health interventions were generally acceptable to LGBTIQ+ young people, and there were some clear themes in aspects of these interventions that determined users' level of interest. Gamification stood out as a component of interventions that tended to be highly regarded by participants. Information presented with brevity and in a relatable way tended to receive greater ratings of acceptability from users and, although infrequent, social aspects of interventions, such as the ability to share experiences with others, were generally highly rated as well. Common concerns raised about the interventions included information being too text heavy, patronizing or contrived, and tasks feeling too laborious or *homework-like*. Importantly, users voiced concerns regarding the targeting of the interventions to LGBTIQ+ people, for example, being targeted at a superficial level, or coming across as stereotypical and alienating in its presentation of LGBTIQ+ people.

Regarding the feasibility of the interventions, measures of engagement and adherence were often not reported by the included studies; however, those that did report levels of use (eg, screen time, clicks, communication with other users) appeared adequate, given the intended scope of the intervention. One study that included an in-person component (collecting rewards earned in the web-based component) reported low user engagement with this feature (5%-27%). Overall, rates of attrition among the included studies were low, with several interventions reporting retention rates of 90% to 100% in their trials [66,69,77]. Notably, however, a trial of one of these interventions in a community setting reported a much lower retention rate of 45.4% [67], indicating that engagement may be lower in reality than controlled trials would suggest. Other studies reported retention rates of 70% to 90%. Two other studies reported notably higher rates of attrition than the others [65,89]. These high rates of attrition did not appear linked to acceptability, with both interventions reporting largely positive responses from participants. There were no notable differences in acceptability and feasibility based on the health outcome interventions.

Discussion

LGBTIQ+ young people have a substantially higher risk of a range of health difficulties than the general population [1], and targeted digital health interventions have the potential to play a crucial role in reducing these disparities [30,33]. The aims of this review were to (1) synthesize the scope of evidence-based, targeted digital health interventions that exist for this population, (2) to identify the overall effectiveness, acceptability, and feasibility of these interventions in this population, and (3) to provide recommendations for their development. The review identified many interventions designed to improve health in LGBTIQ+ young people, and these interventions have shown preliminary effectiveness in producing changes in some health outcomes in this group. Particularly promising evidence was found for the effectiveness of digital health interventions in certain aspects of managing the risk of STIs, notably increasing HIV testing rates, and emerging evidence was also found for reductions in internalizing symptoms and substance use. The review observed a trend that digital health interventions for LGBTIQ+ young people may more consistently effect change in cognitive and affective outcomes than behavioral outcomes, though this was not prescriptive. All of these findings must be considered in light of the preliminary nature of the majority of the studies included in this review and their resultant methodological limitations.

In addition to showing potential for effectiveness, the interventions were found to be generally acceptable and feasible overall. Acceptability appeared closely linked to collaborative intervention design development with LGBTIQ+ young people and the digital modality of delivery. Notably, one study, which had regular check-ins with a clinician, found that participants were in fact deterred by this contact, citing difficulty scheduling and desire to remain discreet [96]. There is currently insufficient evidence to conclude that digital health interventions would be more effective than untargeted or face-to-face interventions in controlled conditions, as most of the studies outlined in this review did not provide such comparisons. However, results such as these suggest that digital health interventions may not need to be more effective than other forms of intervention to be valuable; these interventions would likely engage sections of the LGBTIQ+ youth population who would otherwise be deterred from seeking any support at all. In addition, rates of attrition in the included interventions were lower than those reported in similar interventions for other groups [37,39-41]. When combined with the innately greater potential for dissemination and cost effectiveness that comes with the digital medium, these findings support the premise that digital health interventions may be an important avenue for reducing health disparities in LGBTIQ+ young people in the future.

Future Directions

Overall, the results of this review are therefore promising for the continued development of digital health interventions for LGBTIQ+ young people, and there are some clear paths forward for how this field of research could be developed further. Most of the interventions included in this review have thus far only been evaluated in terms of usability, acceptability, and preliminary efficacy, and due to the preliminary nature of most of these studies, when efficacy was evaluated, aspects of methodology such as randomization and blinding were often lacking in rigor. The findings of this review are therefore consistent with others in the literature that frequently report low quality of evidence associated with digital health interventions [43,45,102,103]. In addition, only 1 study trialed a digital health intervention in a community setting [61], and its rate of user retention was approximately half that of its clinical trial, consistent with previous studies that have found a similar pattern [104]. The generalizability of the findings of these studies to the wider population of LGBTIQ+ young people, therefore, cannot be determined. There is a clear need to build on the emerging evidence base through more rigorous randomized controlled trials and trials in community settings, as this evidence base is crucial for further funding and dissemination. While the evidence thus far is promising, it needs significant development.

Outside of these methodological concerns, there is also notable progress in terms of expanding the scope of the digital health interventions that exist for this population. Most of the interventions identified in this review were directed toward improving the health of young men who identify as gay, bisexual, and queer and largely within the scope of improving sexual health–related concerns alone. Given the high rates of mental and physical health difficulties in LGBTIQ+ people, resources should be directed toward the development of digital health interventions targeting these issues, commensurate with the attention being given to STI and other sexual health–related concerns. The health concerns faced by LBTIQ+ women, trans and gender-diverse people, and people with intersex variations demand attention as well. Trans and gender-diverse people in particular face increased and wide-ranging health difficulties and barriers to health care access, even when compared with other members of the LGBTIQ+ community [25,27,105], and their marked underrepresentation in the interventions included in this review reveals a significant missed opportunity to begin to address these inequalities. While minority subpopulations are typically numerically small, and some argue that developing tailored interventions for such groups may be unnecessary [106], there is also evidence to suggest that minority groups appreciate tailoring of interventions. For example, in a recent study examining the attitudes of transgender and gender-diverse young people toward mental health gaming interventions, participants noted that TGD representation and inclusion of meaningful, specific tailored content was favorable [31].

Limitations

This review is the first to provide a wide overview of digital health interventions for LGBTIQ+ young people, enabling gaps such as these to be highlighted. However, the review is limited in several ways. Owing to the restrictions placed on the degree of human guidance permissible for inclusion in the review, telehealth interventions are notably absent from those discussed here. The review also did not cover interventions for which their development has been documented but has not yet been evaluated in some respects, and the review also did not include studies documenting the effectiveness of nontargeted interventions for LGBTIQ+ young people. Furthermore, the requirement for included studies to be published in the English language may have resulted in a biased sample, potentially excluding reports on interventions published in other languages. Finally, it is possible that limiting the search to studies published after January 1, 2000, may have excluded relevant studies; however, given that the earliest published study identified in the review was not until 10 years later, this is unlikely to have been the case. The review therefore largely covers the scope of interventions that accumulate an evidence base but should not be taken to cover any and all digital health interventions that may benefit the health of LGBTIQ+ youth.

Conclusions

Although not sought out specifically in the process of conducting this review, many protocols have been identified for continued research into the development of digital health interventions [107-112]. The development of digital health interventions for LGBTIQ+ young people is a burgeoning field of study, and we expect the evidence base to advance quickly. Going forward, this advancement should ideally occur across the breadth of health difficulties and inequalities that the entire LGBTIQ+ community faces and with appropriate methodological rigor. Given the number of interventions already targeting risk reduction or management of STIs, rather than developing more, future studies should ideally seek to refine and adapt those that exist for public use, and explore implementation barriers and facilitators to enhance translation. Given the lower scope and evidence base for interventions targeting mental and physical health difficulties, future studies should focus predominantly on expanding the available interventions and evidence base in these domains, particularly in terms of addressing difficulties such as alcohol use and suicide for which no digital health interventions were detected in this review at all. Digital health interventions for LGBTIQ+ young people show the potential to improve health disparities in this population, and the expansion of research along these lines is crucial to realize this potential.

Acknowledgments

This work was funded by the Giorgetta Charity Fund. DG and HM are supported by Research Training Program Stipends funded by the Australian Government. Associate Professor AL is funded by a National Health and Medical Research Council Career Development Fellowship (#1148793). Dr YP is supported by a fellowship from the Giorgetta Charity Fund.

Authors' Contributions

DG developed and executed the search strategy, co-screened the papers, co-assessed their quality, extracted and interpreted the data, and contributed to writing and editing the manuscript. HM co-conducted the screening process, assessed the quality of the papers, and contributed to writing and editing the manuscript. Associate Professor AL contributed to the interpretation of the data and to writing and editing the manuscript; Dr YP conceived and coordinated the study, advised on the development of the search and screening processes, and contributed to writing and editing the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for PsycINFO (Ovid).
[DOCX File , 14 KB - jmir_v22i12e20158_app1.docx]

Multimedia Appendix 2
Summary of digital sexually transmitted infection risk reduction and management interventions for LGBTIQ+ young people.
[DOCX File , 75 KB - jmir_v22i12e20158_app2.docx]

Multimedia Appendix 3
Summary of interventions.
[DOCX File , 66 KB - jmir_v22i12e20158_app3.docx]

References

1. Zeeman L, Sherriff N, Browne K, McGlynn N, Mirandola M, Gios L, Health4LGBTI Network. A review of lesbian, gay, bisexual, trans and intersex (LGBTI) health and healthcare inequalities. Eur J Public Health 2019 Oct 1;29(5):974-980 [FREE Full text] [doi: 10.1093/eurpub/cky226] [Medline: 30380045]
2. Eskin M, Kaynak-Demir H, Demir S. Same-sex sexual orientation, childhood sexual abuse, and suicidal behavior in university students in Turkey. Arch Sex Behav 2005 Apr;34(2):185-195. [doi: 10.1007/s10508-005-1796-8] [Medline: 15803252]
3. Marshal MP, Dietz LJ, Friedman MS, Stall R, Smith HA, McGinley J, et al. Suicidality and depression disparities between sexual minority and heterosexual youth: a meta-analytic review. J Adolesc Health 2011 Aug;49(2):115-123 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.02.005] [Medline: 21783042]
4. Jones T. The needs of students with intersex variations. Sex Educ 2016 Mar 11;16(6):602-618. [doi: 10.1080/14681811.2016.1149808]
5. HIV and Youth. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/hiv/pdf/group/age/youth/cdc-hiv-youth.pdf [accessed 2020-11-16]
6. Marshal MP, Friedman MS, Stall R, King KM, Miles J, Gold MA, et al. Sexual orientation and adolescent substance use: a meta-analysis and methodological review. Addiction 2008 Apr;103(4):546-556 [FREE Full text] [doi: 10.1111/j.1360-0443.2008.02149.x] [Medline: 18339100]
7. Corliss HL, Rosario M, Wypij D, Wylie SA, Frazier AL, Austin SB. Sexual orientation and drug use in a longitudinal cohort study of US adolescents. Addict Behav 2010 May;35(5):517-521 [FREE Full text] [doi: 10.1016/j.addbeh.2009.12.019] [Medline: 20061091]

8. Day JK, Fish JN, Perez-Brumer A, Hatzenbuehler ML, Russell ST. Transgender youth substance use disparities: results from a population-based sample. J Adolesc Health 2017 Dec;61(6):729-735 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.06.024] [Medline: 28942238]
9. Bishop A, Overcash F, McGuire J, Reicks M. Diet and physical activity behaviors among adolescent transgender students: school survey results. J Adolesc Health 2020 Apr;66(4):484-490. [doi: 10.1016/j.jadohealth.2019.10.026] [Medline: 31959401]
10. Saewyc EM, Poon CS, Homma Y, Skay CL. Stigma management? The links between enacted stigma and teen pregnancy trends among gay, lesbian, and bisexual students in British Columbia. Can J Hum Sex 2008;17(3):123-139 [FREE Full text] [Medline: 19293941]
11. Quinn GP, Sanchez JA, Sutton SK, Vadaparampil ST, Nguyen GT, Green BL, et al. Cancer and lesbian, gay, bisexual, transgender/transsexual, and queer/questioning (LGBTQ) populations. CA Cancer J Clin 2015;65(5):384-400 [FREE Full text] [doi: 10.3322/caac.21288] [Medline: 26186412]
12. Case P, Austin SB, Hunter DJ, Manson JE, Malspeis S, Willett WC, et al. Sexual orientation, health risk factors, and physical functioning in the Nurses' Health Study II. J Womens Health (Larchmt) 2004 Nov;13(9):1033-1047. [doi: 10.1089/jwh.2004.13.1033] [Medline: 15665660]
13. Caceres BA, Brody AA, Halkitis PN, Dorsen C, Yu G, Chyun DA. Sexual orientation, health risk factors, and physical functioning in the nurses' health study II. LGBT Health 2018 Jul;5(5):284-294 [FREE Full text] [doi: 10.1089/lgbt.2017.0220] [Medline: 29889585]
14. Caceres BA, Brody AA, Halkitis PN, Dorsen C, Yu G, Chyun DA. Cardiovascular disease risk in sexual minority women (18-59 years old): findings from the national health and nutrition examination survey (2001-2012). Womens Health Issues 2018;28(4):333-341 [FREE Full text] [doi: 10.1016/j.whi.2018.03.004] [Medline: 29661697]
15. Kang M, Bernard D, Booth M, Quine S, Alperstein G, Usherwood T, et al. Access to primary health care for Australian young people: service provider perspectives. Br J Gen Pract 2003 Dec;53(497):947-952 [FREE Full text] [Medline: 14960219]
16. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psych 2010 Dec 30;10(1):-. [doi: 10.1186/1471-244x-10-113]
17. Ray N. Lesbian, Gay, Bisexual and Transgender Youth: an Epidemic of Homelessness. National Gay and Lesbian Task Force Policy Institute and the National Coalition for the Homeless. 2006. URL: https://myhealth.alberta.ca/myhealthrecords#:~:text=You%20need%20to%20sign%20up,License%20or%20Alberta%20Identification%20card [accessed 2020-11-16]
18. Durso L, Gates G. Serving Our Youth: Findings from a National Survey of Service Providers Working with Lesbian, Gay, Bisexual, and Transgender Youth who are Homeless or At Risk of Becoming Homeless. The Williams Institute with True Colors Fund and The Palette Fund. 2012. URL: https://williamsinstitute.law.ucla.edu/publications/serving-our-youth-lgbtq/ [accessed 2020-11-16]
19. Morton MH, Dworsky A, Matjasko JL, Curry SR, Schlueter D, Chávez R, et al. Prevalence and correlates of youth homelessness in the United States. J Adolesc Health 2018 Jan;62(1):14-21 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.10.006] [Medline: 29153445]
20. Sears B, Mallory C. The Williams Institute. 2011. URL: https://williamsinstitute.law.ucla.edu/publications/employ-discrim-effect-lgbt-people/ [accessed 2020-11-16]
21. Takács J. Social Exclusion of Young Lesbian, Gay, Bisexual and Transgender (LGBT) People in Europe. ILGA Europe. 2006. URL: https://www.salto-youth.net/downloads/4-17-948/ReportSocialExclusionIGLYOilga.pdf [accessed 2020-11-16]
22. Brown A, Rice SM, Rickwood DJ, Parker AG. Systematic review of barriers and facilitators to accessing and engaging with mental health care among at-risk young people. Asia Pac Psychiatry 2016 Mar;8(1):3-22. [doi: 10.1111/appy.12199] [Medline: 26238088]
23. Sanchez NF, Sanchez JP, Danoff A. Health care utilization, barriers to care, and hormone usage among male-to-female transgender persons in New York City. Am J Public Health 2009 Apr;99(4):713-719. [doi: 10.2105/AJPH.2007.132035] [Medline: 19150911]
24. Mayer KH, Bradford JB, Makadon HJ, Stall R, Goldhammer H, Landers S. Sexual and gender minority health: what we know and what needs to be done. Am J Public Health 2008 Jun;98(6):989-995. [doi: 10.2105/AJPH.2007.127811] [Medline: 18445789]
25. Safer JD, Coleman E, Feldman J, Garofalo R, Hembree W, Radix A, et al. Barriers to healthcare for transgender individuals. Curr Opin Endocrinol Diabetes Obes 2016;23(2):168-171. [doi: 10.1097/med.0000000000000227]
26. McIntyre J, Daley A, Rutherford K, Ross LE. Systems-level barriers in accessing supportive mental health services for sexual and gender minorities: insights from the provider's perspective. Can J Commun Ment Health 2011 Sep;30(2):173-186. [doi: 10.7870/cjcmh-2011-0023]
27. Macapagal K, Bhatia R, Greene GJ. Differences in healthcare access, use, and experiences within a community sample of racially diverse lesbian, gay, bisexual, transgender, and questioning emerging adults. LGBT Health 2016 Dec;3(6):434-442 [FREE Full text] [doi: 10.1089/lgbt.2015.0124] [Medline: 27726496]

28. Kates J, Ranji U, Beamesderfer A, Salganicoff A, Dawson L. Health and Access to Care and Coverage for Lesbian, Gay, Bisexual,and Transgender (LGBT) Individuals in the U.S. Kaiser Family Foundation. 2018. URL: https://www.kff.org/racial-equity-and-health-policy/issue-brief/health-and-access-to-care-and-coverage-for-lesbian-gay-bisexual-and-transgender-individuals-in-the-u-s/ [accessed 2020-11-16]
29. Ward B, Dahlhammer J, Galinsky A, Joestl S. Sexual Orientation and Health Among US Adults: National Health Interview Survey. National Health Statistics Reports. 2014. URL: https://www.cdc.gov/nchs/nhis/sexual_orientation/statistics.htm [accessed 2020-11-16]
30. Perry Y, Strauss P, Lin A. Online interventions for the mental health needs of trans and gender diverse young people. Lancet Psychiatry 2018 Feb;5(2):e6. [doi: 10.1016/S2215-0366(18)30017-8] [Medline: 29413141]
31. Strauss P, Morgan H, Wright Toussaint D, Lin A, Winter S, Perry Y. Trans and gender diverse young people's attitudes towards game-based digital mental health interventions: A qualitative investigation. Internet Interv 2019 Dec;18:100280 [FREE Full text] [doi: 10.1016/j.invent.2019.100280] [Medline: 31890628]
32. Lucassen MF, Hatcher S, Stasiak K, Fleming T, Shepherd M, Merry SN. The views of lesbian, gay and bisexual youth regarding computerised self-help for depression: An exploratory study. Adv Mental Health 2014 Dec 17;12(1):22-33. [doi: 10.5172/jamh.2013.12.1.22]
33. Bowen D, Jabson J, Kamen C. mHealth: an avenue for promoting health among sexual and gender minority populations? Mhealth 2016;2:36 [FREE Full text] [doi: 10.21037/mhealth.2016.09.01] [Medline: 28293609]
34. Aguilera A. Digital technology and mental health interventions: opportunities and challenges. Arbor 2015 May 12;191(771):a210. [doi: 10.3989/arbor.2015.771n1012]
35. Berry N, Lobban F, Bucci S. A qualitative exploration of service user views about using digital health interventions for self-management in severe mental health problems. BMC Psychiatry 2019 Jan 21;19(1):35 [FREE Full text] [doi: 10.1186/s12888-018-1979-1] [Medline: 30665384]
36. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
37. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
38. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
39. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an ehealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
40. Schubart JR, Stuckey HL, Ganesh A, Sciamanna CN. Chronic health conditions and internet behavioral interventions. Comput Info Nurs 2011;29(2):81-92. [doi: 10.1097/ncn.0b013e3182065eed]
41. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
42. Kiluk BD, Sugarman DE, Nich C, Gibbons CJ, Martino S, Rounsaville BJ, et al. A methodological analysis of randomized clinical trials of computer-assisted therapies for psychiatric disorders: toward improved standards for an emerging field. Am J Psychiatry 2011 Aug;168(8):790-799 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10101443] [Medline: 21536689]
43. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
44. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
45. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
46. Strauss P, Cook A, Winter S, Watson V, Wright-Toussaint D, Lin A. Trans-pathways: the Mental Health Experiences and Care Pathways of Trans Young People-summary of Results. Telethon Kids Institute. 2017. URL: https://www.telethonkids.org.au/projects/trans-pathways/ [accessed 2020-11-16]
47. Craig SL, McInroy L. You can form a part of yourself online: the influence of new media on identity development and coming out for LGBTQ youth. J Gay Lesbian Mental Health 2014 Jan 30;18(1):95-109. [doi: 10.1080/19359705.2013.777007]
48. McDermott E, Hughes E, Rawlings V. Queer Futures. Understanding Lesbian, Gay, Bisexual and Trans (LGBT) Adolescents' Suicide, Self-harm and Help-seeking Behaviour. Department of Health Policy Research Programme Project. 2016. URL: http://www.queerfutures.co.uk/wp-content/uploads/2016/06/Queer-Futures-Final-Report.pdf [accessed 2020-11-16]
49. Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, et al. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. PLoS ONE 2017 Aug 16;12(8):e0181722. [doi: 10.1371/journal.pone.0181722]

50. Mustanski B, Lyons T, Garcia SC. Internet use and sexual health of young men who have sex with men: a mixed-methods study. Arch Sex Behav 2011 Apr;40(2):289-300 [FREE Full text] [doi: 10.1007/s10508-009-9596-1] [Medline: 20182787]
51. Rozbroj T, Lyons A, Pitts M, Mitchell A, Christensen H. Assessing the applicability of e-therapies for depression, anxiety, and other mood disorders among lesbians and gay men: analysis of 24 web- and mobile phone-based self-help interventions. J Med Internet Res 2014 Jul 3;16(7):e166 [FREE Full text] [doi: 10.2196/jmir.3529] [Medline: 24996000]
52. Steinke J, Root-Bowman M, Estabrook S, Levine DS, Kantor LM. Meeting the needs of sexual and gender minority youth: formative research on potential digital health interventions. J Adolesc Health 2017 May;60(5):541-548 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.11.023] [Medline: 28108088]
53. Lucassen M, Samra R, Iacovides I, Fleming T, Shepherd M, Stasiak K, et al. How LGBT+ young people use the internet in relation to their mental health and envisage the use of e-therapy: exploratory study. JMIR Serious Games 2018 Dec 21;6(4):e11249 [FREE Full text] [doi: 10.2196/11249] [Medline: 30578194]
54. Knight R, Karamouzian M, Salway T, Gilbert M, Shoveller J. Online interventions to address HIV and other sexually transmitted and blood-borne infections among young gay, bisexual and other men who have sex with men: a systematic review. J Int AIDS Soc 2017 Nov;20(3):- [FREE Full text] [doi: 10.1002/jia2.25017] [Medline: 29091340]
55. Hobaica S, Alman A, Jackowich S, Kwon P. Empirically based psychological interventions with sexual minority youth: a systematic review. Psychol Sex Orientat Gend Divers 2018 Sep;5(3):313-323. [doi: 10.1037/sgd0000275]
56. van der Pol-Harney E, McAloon J. Psychosocial interventions for mental illness among LGBTQIA youth: a prisma-based systematic review. Adolescent Res Rev 2018 Aug 2;4(2):149-168. [doi: 10.1007/s40894-018-0090-7]
57. Coulter RW, Egan JE, Kinsky S, Friedman MR, Eckstrand KL, Frankeberger J, et al. Mental health, drug, and violence interventions for sexual/gender minorities: a systematic review. Pediatrics 2019 Sep;144(3):- [FREE Full text] [doi: 10.1542/peds.2018-3367] [Medline: 31427462]
58. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
59. Howarth A, Quesada J, Silva J, Judycki S, Mills PR. The impact of digital health interventions on health-related outcomes in the workplace: a systematic review. Digit Health 2018;4:2055207618770861 [FREE Full text] [doi: 10.1177/2055207618770861] [Medline: 29942631]
60. Mahood Q, van Eerd D, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. Res Synth Methods 2014 Sep;5(3):221-234. [doi: 10.1002/jrsm.1106] [Medline: 26052848]
61. Gilbey D, Mahfouda S, Ohan J, Lin A, Perry Y. Trajectories of mental health difficulties in young people who are attracted to the same gender: a systematic review. Adolescent Res Rev 2019 Aug 9;5(3):281-293. [doi: 10.1007/s40894-019-00128-8]
62. Lee JG, Ylioja T, Lackey M. Identifying lesbian, gay, bisexual, and transgender search terminology: a systematic review of health systematic reviews. PLoS One 2016;11(5):e0156210 [FREE Full text] [doi: 10.1371/journal.pone.0156210] [Medline: 27219460]
63. Hong Q, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Mixed Methods Appraisal Tool (MMAT), Version 2018. IC Canadian Intellectual Property Office, Industry Canada. 2018. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf [accessed 2020-11-16]
64. Mustanski B, Greene GJ, Ryan D, Whitton SW. Feasibility, acceptability, and initial efficacy of an online sexual health promotion program for LGBT youth: the Queer Sex Ed intervention. J Sex Res 2015;52(2):220-230. [doi: 10.1080/00224499.2013.867924] [Medline: 24588408]
65. Martin S. Developing and Evaluating Queervibe: an Online Intervention to Empower Trans and Non-binary Youth. Doctoral Dissertation, Anglia Ruskin University. 2019. URL: https://arro.anglia.ac.uk/704647/ [accessed 2020-11-16]
66. Mitchell JT, LeGrand S, Hightow-Weidman LB, McKellar MS, Kashuba AD, Cottrell M, et al. Smartphone-based contingency management intervention to improve pre-exposure prophylaxis adherence: pilot trial. JMIR Mhealth Uhealth 2018 Sep 10;6(9):e10456 [FREE Full text] [doi: 10.2196/10456] [Medline: 30201601]
67. Greene GJ, Madkins K, Andrews K, Dispenza J, Mustanski B. Implementation and evaluation of the keep it up! Online HIV prevention intervention in a community-based setting. AIDS Educ Prev 2016 Jun;28(3):231-245. [doi: 10.1521/aeap.2016.28.3.231] [Medline: 27244191]
68. Madkins K, Moskowitz DA, Moran K, Dellucci TV, Mustanski B. Measuring acceptability and engagement of the keep it up! Internet-based HIV prevention randomized controlled trial for young men who have sex with men. AIDS Educ Prev 2019 Aug;31(4):287-305 [FREE Full text] [doi: 10.1521/aeap.2019.31.4.287] [Medline: 31361516]
69. Mustanski B, Garofalo R, Monahan C, Gratzer B, Andrews R. Feasibility, acceptability, and preliminary efficacy of an online HIV prevention program for diverse young men who have sex with men: the keep it up! intervention. AIDS Behav 2013 Nov;17(9):2999-3012 [FREE Full text] [doi: 10.1007/s10461-013-0507-z] [Medline: 23673793]
70. Mustanski B, Parsons JT, Sullivan PS, Madkins K, Rosenberg E, Swann G. Biomedical and behavioral outcomes of keep it up!: an ehealth HIV prevention program RCT. Am J Prev Med 2018 Aug;55(2):151-158 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.026] [Medline: 29937115]

71. Christensen JL, Miller LC, Appleby PR, Corsbie-Massay C, Godoy CG, Marsella SC, et al. Reducing shame in a game that predicts HIV risk reduction for young adult men who have sex with men: a randomized trial delivered nationally over the web. J Int AIDS Soc 2013 Nov 13;16:18716. [doi: 10.7448/ias.16.3.18716]
72. Hightow-Weidman LB, Fowler B, Kibe J, McCoy R, Pike E, Calabria M, et al. HealthMpowerment.org: development of a theory-based HIV/STI website for young black MSM. AIDS Educ Prev 2011 Feb;23(1):1-12 [FREE Full text] [doi: 10.1521/aeap.2011.23.1.1] [Medline: 21341956]
73. Hightow-Weidman LB, LeGrand S, Muessig KE, Simmons RA, Soni K, Choi SK, et al. A randomized trial of an online risk reduction intervention for young black MSM. AIDS Behav 2019 May;23(5):1166-1177 [FREE Full text] [doi: 10.1007/s10461-018-2289-9] [Medline: 30269231]
74. Hightow-Weidman LB, Pike E, Fowler B, Matthews DM, Kibe J, McCoy R, et al. HealthMpowerment.org: feasibility and acceptability of delivering an internet intervention to young Black men who have sex with men. AIDS Care 2012;24(7):910-920 [FREE Full text] [doi: 10.1080/09540121.2011.647677] [Medline: 22272759]
75. Ybarra ML, DuBois LZ, Parsons JT, Prescott TL, Mustanski B. Online focus groups as an HIV prevention program for gay, bisexual, and queer adolescent males. AIDS Educ Prev 2014 Dec;26(6):554-564 [FREE Full text] [doi: 10.1521/aeap.2014.26.6.554] [Medline: 25490735]
76. Ybarra ML, Liu W, Prescott TL, Phillips G, Mustanski B. The effect of a text messaging based HIV prevention program on sexual minority male youths: a national evaluation of information, motivation and behavioral skills in a randomized controlled trial of guy2guy. AIDS Behav 2018 Oct;22(10):3335-3344 [FREE Full text] [doi: 10.1007/s10461-018-2118-1] [Medline: 29696403]
77. Ybarra ML, Prescott T, Mustanski B, Parsons J, Bull SS. Feasibility, acceptability, and process indicators for guy2guy, an mhealth HIV prevention program for sexual minority adolescent boys. J Adolesc Health 2019 Sep;65(3):417-422 [FREE Full text] [doi: 10.1016/j.jadohealth.2019.04.025] [Medline: 31277991]
78. Ybarra ML, Prescott TL, Philips GL, Bull SS, Parsons JT, Mustanski B. Iteratively developing an mhealth HIV prevention program for sexual minority adolescent men. AIDS Behav 2016 Jun;20(6):1157-1172 [FREE Full text] [doi: 10.1007/s10461-015-1146-3] [Medline: 26238038]
79. Ybarra ML, Prescott TL, Phillips GL, Bull SS, Parsons JT, Mustanski B. Pilot RCT results of an mhealth hiv prevention program for sexual minority male adolescents. Pediatrics 2017 Jul;140(1):- [FREE Full text] [doi: 10.1542/peds.2016-2999] [Medline: 28659456]
80. Solorio R, Norton-Shelpuk P, Forehand M, Montaño D, Stern J, Aguirre J, et al. Tu amigo pepe: evaluation of a multi-media marketing campaign that targets young Latino immigrant MSM with HIV testing messages. AIDS Behav 2016 Sep;20(9):1973-1988. [doi: 10.1007/s10461-015-1277-6] [Medline: 26850101]
81. Cho H, Powell D, Pichon A, Thai J, Bruce J, Kuhns LM, et al. A mobile health intervention for HIV prevention among racially and ethnically diverse young men: usability evaluation. JMIR Mhealth Uhealth 2018 Sep 7;6(9):e11450 [FREE Full text] [doi: 10.2196/11450] [Medline: 30194060]
82. Muessig K, Knudtson KA, Soni K, Larsen MA, Traum D, Dong W, et al. 'I didn't tell you sooner because I didn't know how to handle it myself.' Developing a virtual reality program to support HIV-status disclosure decisions. Digit Cult Educ 2018;10:22-48 [FREE Full text] [Medline: 30123342]
83. Bauermeister JA, Tingler RC, Demers M, Connochie D, Gillard G, Shaver J, et al. Acceptability and preliminary efficacy of an online HIV prevention intervention for single young men who have sex with men seeking partners online: the mydex project. AIDS Behav 2019 Nov;23(11):3064-3077. [doi: 10.1007/s10461-019-02426-7] [Medline: 30762190]
84. Goldstone S. LBO-1.6 efficacy of the quadrivalent HPV vaccine to prevent anal intraepithelial neoplasia among young men who have sex with men. Sex Transmit Infect 2011 Jul 10;87(Suppl 1):A352. [doi: 10.1136/sextrans-2011-050119.6]
85. McRee A, Shoben A, Bauermeister JA, Katz ML, Paskett ED, Reiter PL. Outsmart HPV: acceptability and short-term effects of a web-based HPV vaccination intervention for young adult gay and bisexual men. Vaccine 2018 Dec 18;36(52):8158-8164 [FREE Full text] [doi: 10.1016/j.vaccine.2018.01.009] [Medline: 29331245]
86. Bauermeister JA, Pingel ES, Jadwin-Cakmak L, Harper GW, Horvath K, Weiss G, et al. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS Behav 2015 Oct;19(10):1860-1874 [FREE Full text] [doi: 10.1007/s10461-015-1009-y] [Medline: 25638038]
87. Horvath KJ, Bauermeister JA. Ehealth literacy and intervention tailoring impacts the acceptability of a HIV/STI testing intervention and sexual decision making among young gay and bisexual men. AIDS Educ Prev 2017 Feb;29(1):14-23 [FREE Full text] [doi: 10.1521/aeap.2017.29.1.14] [Medline: 28195779]
88. Washington TA, Applewhite S, Glenn W. Using facebook as a platform to direct young black men who have sex with men to a video-based HIV testing intervention: a feasibility study. Urban Soc Work 2017;1(1):36-52 [FREE Full text] [doi: 10.1891/2474-8684.1.1.36] [Medline: 29276800]
89. McCoy SI, Buzdugan R, Grimball R, Natoli L, Mejia CM, Klausner JD, et al. pilot study results of an intervention using gamification to increase HIV screening among young men who have sex with men in California. Mhealth 2018;4:40 [FREE Full text] [doi: 10.21037/mhealth.2018.09.04] [Medline: 30363751]

90. LeGrand S, Muessig KE, McNulty T, Soni K, Knudtson K, Lemann A, et al. Epic allies: development of a gaming app to improve antiretroviral therapy adherence among young HIV-positive men who have sex with men. JMIR Serious Games 2016 May 13;4(1):e6 [FREE Full text] [doi: 10.2196/games.5687] [Medline: 27178752]
91. Hightow-Weidman L, Muessig K, Knudtson K, Srivatsa M, Lawrence E, LeGrand S, et al. A gamified smartphone app to support engagement in care and medication adherence for HIV-positive young men who have sex with men (allyquest): development and pilot study. JMIR Public Health Surveill 2018 Apr 30;4(2):e34 [FREE Full text] [doi: 10.2196/publichealth.8923] [Medline: 29712626]
92. Schwinn TM, Thom B, Schinke SP, Hopkins J. Preventing drug use among sexual-minority youths: findings from a tailored, web-based intervention. J Adolesc Health 2015 May;56(5):571-573. [doi: 10.1016/j.jadohealth.2014.12.015] [Medline: 25744209]
93. Lucassen M, Merry S, Hatcher S. That's [not] so gay: computerised cognitive behavioural therapy for depressive symptoms in sexual minority youth. Neuropsychiatrie de l'Enfance et de l'Adolescence 2012 Jul;60(5):S141. [doi: 10.1016/j.neurenf.2012.04.116]
94. Lucassen MF, Hatcher S, Fleming TM, Stasiak K, Shepherd MJ, Merry SN. A qualitative study of sexual minority young people's experiences of computerised therapy for depression. Australas Psychiatry 2015 Jun;23(3):268-273. [doi: 10.1177/1039856215579542] [Medline: 25881962]
95. Lucassen MF, Merry SN, Hatcher S, Frampton CM. Rainbow sparx: a novel approach to addressing depression in sexual minority youth. Cogn Behav Pract 2015 May;22(2):203-216. [doi: 10.1016/j.cbpra.2013.12.008]
96. Fleming JB, Hill YN, Burns MN. Usability of a culturally informed mhealth intervention for symptoms of anxiety and depression: feedback from young sexual minority men. JMIR Hum Factors 2017 Aug 25;4(3):e22 [FREE Full text] [doi: 10.2196/humanfactors.7392] [Medline: 28842389]
97. Pachankis JE, Goldfried MR. Expressive writing for gay-related stress: psychosocial benefits and mechanisms underlying improvement. J Consult Clin Psychol 2010 Feb;78(1):98-110. [doi: 10.1037/a0017580] [Medline: 20099955]
98. Vogel EA, Belohlavek A, Prochaska JJ, Ramo DE. Development and acceptability testing of a Facebook smoking cessation intervention for sexual and gender minority young adults. Internet Interv 2019 Mar;15:87-92 [FREE Full text] [doi: 10.1016/j.invent.2019.01.002] [Medline: 30792958]
99. Vogel EA, Thrul J, Humfleet GL, Delucchi KL, Ramo DE. Smoking cessation intervention trial outcomes for sexual and gender minority young adults. Health Psychol 2019 Jan;38(1):12-20 [FREE Full text] [doi: 10.1037/hea0000698] [Medline: 30489104]
100. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. Br Med J 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
101. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Increasing human papillomavirus vaccination among young gay and bisexual men: a randomized pilot trial of the outsmart HPV intervention. LGBT Health 2018 Jul;5(5):325-329 [FREE Full text] [doi: 10.1089/lgbt.2018.0059] [Medline: 29979642]
102. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):89-103. [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
103. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
104. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
105. Su D, Irwin JA, Fisher C, Ramos A, Kelley M, Mendoza DA, et al. Mental health disparities within the LGBT population: a comparison between transgender and nontransgender individuals. Transgend Health 2016;1(1):12-20 [FREE Full text] [doi: 10.1089/trgh.2015.0001] [Medline: 29159294]
106. Pachankis JE. The scientific pursuit of sexual and gender minority mental health treatments: toward evidence-based affirmative practice. Am Psychol 2018 Dec;73(9):1207-1219 [FREE Full text] [doi: 10.1037/amp0000357] [Medline: 30525805]
107. Koblin B, Hirshfield S, Chiasson MA, Wilton L, Usher D, Nandi V, et al. Intervention to match young black men and transwomen who have sex with men or transwomen to HIV testing options (all about me): protocol for a randomized controlled trial. JMIR Res Protoc 2017 Dec 19;6(12):e254 [FREE Full text] [doi: 10.2196/resprot.8856] [Medline: 29258976]
108. LeGrand S, Muessig KE, Platt A, Soni K, Egger JR, Nwoko N, et al. Epic allies, a gamified mobile phone app to improve engagement in care, antiretroviral uptake, and adherence among young men who have sex with men and young transgender women who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Apr 5;7(4):e94 [FREE Full text] [doi: 10.2196/resprot.8811] [Medline: 29622527]
109. Bauermeister JA, Golinkoff JM, Horvath KJ, Hightow-Weidman LB, Sullivan PS, Stephenson R. A multilevel tailored web app-based intervention for linking young men who have sex with men to quality care (get connected): protocol for a

randomized controlled trial. JMIR Res Protoc 2018 Aug 2;7(8):e10444 [FREE Full text] [doi: 10.2196/10444] [Medline: 30072358]
110. Coulter RW, Sang JM, Louth-Marquez W, Henderson ER, Espelage D, Hunter SC, et al. Pilot testing the feasibility of a game intervention aimed at improving help seeking and coping among sexual and gender minority youth: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Feb 15;8(2):e12164 [FREE Full text] [doi: 10.2196/12164] [Medline: 30767903]
111. Kuhns LM, Garofalo R, Hidalgo M, Hirshfield S, Pearson C, Bruce J, et al. A randomized controlled efficacy trial of an mHealth HIV prevention intervention for sexual minority young men: MyPEEPS mobile study protocol. BMC Public Health 2020 Jan 15;20(1):65 [FREE Full text] [doi: 10.1186/s12889-020-8180-4] [Medline: 31941475]
112. Reiter PL, Gower AL, Kiss DE, Malone MA, Katz ML, Bauermeister JA, et al. A web-based human papillomavirus vaccination intervention for young gay, bisexual, and other men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Feb 24;9(2):e16294 [FREE Full text] [doi: 10.2196/16294] [Medline: 32130192]

Abbreviations

HPV: human papillomavirus
LGBTIQ+: lesbian, gay, bisexual, transgender, intersex or queer
MMAT: mixed methods appraisal tool
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: Prospective Register of Systematic Reviews
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 12.05.20; peer-reviewed by M Lucassen, M Embrett; comments to author 16.06.20; revised version received 30.07.20; accepted 26.10.20; published 03.12.20.

Please cite as:
Gilbey D, Morgan H, Lin A, Perry Y
Effectiveness, Acceptability, and Feasibility of Digital Health Interventions for LGBTIQ+ Young People: Systematic Review
J Med Internet Res 2020;22(12):e20158
URL: https://www.jmir.org/2020/12/e20158
doi:10.2196/20158
PMID:33270039

Review

Decision Support Capabilities of Telemedicine in Emergency Prehospital Care: Systematic Review

Yesul Kim[1,2,3], BA, PhD; Christopher Groombridge[1,2,3], MBBS, MSc, FACEM; Lorena Romero[4], BA, MBIT; Steven Clare[1,3], BMedSc; Mark Christopher Fitzgerald[1,2,3], MBBS, MD, FACEM

[1]National Trauma Research Institute, Melbourne, Australia
[2]Monash University, Melbourne, Australia
[3]Trauma Services, Alfred Health, Melbourne, Australia
[4]The Ian Potter Library, Alfred Health, Melbourne, Australia

Corresponding Author:
Yesul Kim, BA, PhD
National Trauma Research Institute
89 Commercial Road
Melbourne, 3004
Australia
Phone: 61 03 98768855
Email: yen.kim@alfred.org.au

Abstract

Background: Telemedicine offers a unique opportunity to improve coordination and administration for urgent patient care remotely. In an emergency setting, it has been used to support first responders by providing telephone or video consultation with specialists at hospitals and through the exchange of prehospital patient information. This technological solution is evolving rapidly, yet there is a concern that it is being implemented without a demonstrated clinical need and effectiveness as well as without a thorough economic evaluation.

Objective: Our objective is to systematically review whether the clinical outcomes achieved, as reported in the literature, favor telemedicine decision support for medical interventions during prehospital care.

Methods: This systematic review included peer-reviewed journal articles. Searches of 7 databases and relevant reviews were conducted. Eligibility criteria consisted of studies that covered telemedicine as data- and information-sharing and two-way teleconsultation platforms, with the objective of supporting medical decisions (eg, diagnosis, treatment, and receiving hospital decision) in a prehospital emergency setting. Simulation studies and studies that included pediatric populations were excluded. The procedures in this review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement. The Risk Of Bias In Non-randomised Studies–of Interventions (ROBINS-I) tool was used for the assessment of risk of bias. The results were synthesized based on predefined aspects of medical decisions that are made in a prehospital setting, which include diagnostic decision support, receiving facility decisions, and medical directions for treatment. All data extractions were done by at least two reviewers independently.

Results: Out of 42 full-text reviews, 7 were found eligible. Diagnostic support and medical direction and decision for treatments were often reported. A key finding of this review was the high agreement between prehospital diagnoses via telemedicine and final in-hospital diagnoses, as supported by quantitative evidence. However, a majority of the articles described the clinical value of having access to remote experts without robust quantitative data. Most telemedicine solutions were evaluated within a feasibility or short-term preliminary study. In general, the results were positive for telemedicine use; however, biases, due to preintervention confounding factors and a lack of documentation on quality assurance and protocol for telemedicine activation, make it difficult to determine the direct effect on patient outcomes.

Conclusions: The information-sharing capacity of telemedicine enables access to remote experts to support medical decision making on scene or in prolonged field care. The influence of human and technology factors on patient care is poorly understood and documented.

(J Med Internet Res 2020;22(12):e18959) doi:10.2196/18959

KEYWORDS

telemedicine; decision support; emergency; resuscitation

Introduction

Effective communication has a central role in emergency care and trauma management [1-3]. Team coordination has been enhanced by information technology, which influences the way we work and interact. As an example, telephone consultation requests for remote experts are often made for complex patient scenarios between Level 1 trauma centers and rural hospitals [4]. The advent of portable devices, allowing real-time data and image transfer, has been shown to assist with making diagnoses in stroke and burns [5,6]. Heads-up-display (HUD) devices are being trialed in neurosurgical and orthopedic theaters for remote surgical guidance and vital signs monitoring [7,8]. Similarly, greater network coverage and processing and data speeds have increased the utility of smartphones and wireless devices for prehospital emergency medicine. In recent simulation and pilot studies, remote experts were able to support paramedics, or emergency medical technicians (EMTs), based on relayed data via telemedicine [9-11]. In this review, the term EMT will be used broadly to include all first responders providing medical care for the injured and/or ill.

Telemedicine is defined by the World Health Organization as the use of "Modern information and communication technologies (ICTs), such as computers, the Internet, and cell phones...to improve patient outcomes by increasing access to care and medical information" [12].

For the purpose of this review, the term telemedicine will cover ICT-facilitated two-way teleconsultation and data sharing between prehospital and hospital virtual hubs. Although various terms and methods define telemedicine, the overarching purpose of facilitating prehospital patient care and communication to the receiving hospital remains consistent across existing trials.

Prehospital resuscitation is delivered in a complex environment where responders make decisions under short time frames and with limited information. Factors such as a patient's illness severity, ambient noise levels, fatigue, and stress can hinder decision making [13]. Various checklists and protocols have been designed, targeting out-of-hospital decision making. Over the years, computerized clinical decision support systems have also been implemented for in-hospital critical care and for disaster or combat and casualty decision making to provide teams with safety alerts, protocols, and diagnostic support for trauma management [14-17]. However, many fail to consider unstructured situations, technology resistance, and logistical support issues that may hinder reliable outcomes [18,19].

Research is limited on the decision-making mechanisms EMTs use at individual and organizational levels. It is unknown whether the currently available training programs and checklists mitigate human factors or environmental stressors that impact one's decision-making processes [13]. A small number of studies investigating the decision-making processes of first responders show that they make use of intuitive reasoning, whereby their past experiences and first impressions on scene make the most impact, rather than an application of specific decision aids [20,21]. Importantly, information that cannot be assessed rapidly by visual or auditory scan (eg, blood pressure, heart rate, respiratory rate, and nonobvious anatomic injuries) is generally not part of a primary triage decision-making process [21]. For complex scenarios, a more deductive analytic method is used, with greater attention given to the available data [21].

In addition to the complexity of scenarios, prehospital care providers are now faced with making decisions over different patient care options. EMTs vary in the level of care they can provide [22]. Around the world, many services focus on rapid transport to hospital, thereby de-emphasizing interventions in the field, whereas in certain locations, specialized medical teams are able to undertake highly advanced interventions, such as extracorporeal membrane oxygenation [23,24]. Therefore, decisions over the course of prehospital care are made in the context of organizational constraints, culture, patient acuity, and evolving demands of EMTs.

In its simplest form, telephone consultation between EMTs and remote experts has been recognized as a beneficial tool [25,26], but the impact of telemedicine on clinical practice is poorly described. Understanding the medical decisions that are associated with telemedicine use as an intervention may have meaningful implications for training, protocols, and operations. As a first step, this systematic review aims to examine the evidence for telemedicine use with medical decisions made during prehospital care. Secondly, this review aims to examine whether telemedicine use contributes to decisions being made and whether it is associated with patient benefit.

Methods

The methods in this review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [27].

Data Sources and Searches

The search strategy aimed to find published studies in 7 databases: MEDLINE, Embase, BIOSIS Previews, Emcare, PsycINFO, Scopus, and Web of Science. After an initial search for articles in MEDLINE and Embase, an analysis of the text words contained in the titles, abstracts, and index terms used to describe these articles was conducted. A second search using all identified keyword terms was then undertaken from database inception to March 29, 2019, across all 7 databases. Studies published in English were considered for inclusion in this review. The final searches were based on the MEDLINE search strategy (see Multimedia Appendix 1). Searches were adapted as appropriate to the specifications of all databases. Hand searching and reference checking of citations and reference lists were also undertaken. Authors of relevant studies were contacted if insufficient data were published.

Study Selection

Articles on the implementation of telemedicine comprised of bidirectional communication via any device, with or without image- or data-sharing capabilities, in a prehospital setting for emergency and/or trauma cases were of interest. The following selection criteria were applied: (1) Population: patients attended by EMTs, (2) Intervention: telemedicine, (3) Outcomes: medical decisions made (eg, diagnosis and treatment), receiving facility decision, and any other clinical judgment, and (4) Setting: prehospital emergency.

The exclusion criteria were as follows: (1) the study included a pediatric population, (2) the study included a nonemergency setting, (3) the paper was only available as an abstract or poster, (4) the study included simulated trials, (5) the paper was a case study, editorial, dissertation, protocol, or review, and (6) the study was conducted purely for evaluating the feasibility and effectiveness of a technology. Studies including the pediatric population were excluded from this systematic review due to the incident rates associated with transportation via emergency services. Emergency medical transport use for pediatric cases has been reported at around 5%-10% [28,29]. Furthermore, medical interventions on children may require consent from their parents. Given the differences in incidence levels and medical interventions, this study focused on the telemedicine implementation for the adult population only.

Simulation trials have previously been included in systematic reviews on telemedicine applications in a prehospital setting [30-33]. This systematic review intended to examine telemedicine use in real-life scenarios to better understand its clinical value and limitations, as it is yet to be embedded into widespread routine use. We included studies that did not have a comparator or control group. Although this limited the evidence examined, it provides a comprehensive overview of telemedicine outcomes under actual emergency circumstances.

Data Extraction

First, one reviewer (YK) independently evaluated the titles and abstracts of all records identified in the initial databases search. Two reviewers (YK and CG) then assessed the eligible full-text articles and, if necessary, discussed their suitability. Additionally, references from reviews and journal articles were screened by one of the reviewers. Disagreements on questions of eligibility were resolved through discussion and none of the articles required an escalation to a third reviewer (MF).

Once a decision was reached, in line with the inclusion criteria, the selected full texts were reviewed by YK and CG. Data were extracted regarding the medical emergency category, the telemedicine platform, and each platform's associated decision outcomes. Any medical decisions or clinical judgments made as an outcome of the telemedicine intervention by the journal authors were considered. Technical performance, with regard to device failure and network issues, was noted if it interfered with the delivery of patient care. Study characteristics were also obtained, including sample size, trial location, and study design.

Risk-of-Bias Assessment

After individual assessments, two reviewers discussed and agreed upon the quality assessment of each article. In order to cater to the heterogeneity in study design, the Risk Of Bias In Non-randomised Studies–of Interventions (ROBINS-I) tool [34] was used to evaluate the methodological rigor in 6 studies. This risk assessment tool was deemed appropriate, as it is particularly concerned with evaluating the effectiveness or safety of an intervention from studies that did not use randomization in allocation.

This tool covers 7 domains of bias that can be introduced from using nonrandomization. Two preintervention biases are particularly of concern in nonrandomized studies. Confounding factors considered for telemedicine intervention were the experience level of EMTs, injury severity for all emergencies, and intervention awareness. Failing to control for such confounders may reduce the comparability between the intervention and control groups. Selection bias was considered a risk whenever eligible participants were excluded in a way that could lead to an association between telemedicine and outcomes. An example of a confounder is if no explanation was provided for how telemedicine was implemented for the study group as compared to the controls during the trial period. Classification bias of interventions at the time of the study was related to the information about the delivery and structure of telemedicine, patient medical records, and any organizational records obtained. Problems of recall bias and subjective opinions can be avoided if voice communication was recorded and assessed by persons not involved in resuscitation and if the medical records assessed were a part of routine care prior to the commencement of the study.

Biases can also arise when there are differences in the care provided, between the study and control groups, due to the awareness of expectations or outcomes of the study; most studies cannot be blinded when using telemedicine. Importantly for telemedicine, performance bias and adherence may be factors that were of interest under risk assessment. Biases during the postintervention phase relate to data handling, particularly the measurement of outcomes; how the results were reported; and how missing data were handled. In particular, we noted whether there was an outcome assessment measure that was applied across cases and whether the outcome could easily be influenced by the assessor being aware of telemedicine being implemented.

One of the included studies was a review of cases and did not have enough information to be assessed by ROBINS-I. This particular study was assessed based on confounders, measurement of outcomes, missing data, and reporting biases. The two investigators (YK and CG) discussed the possible biases and the evidence in support of the decisions.

Data Analysis

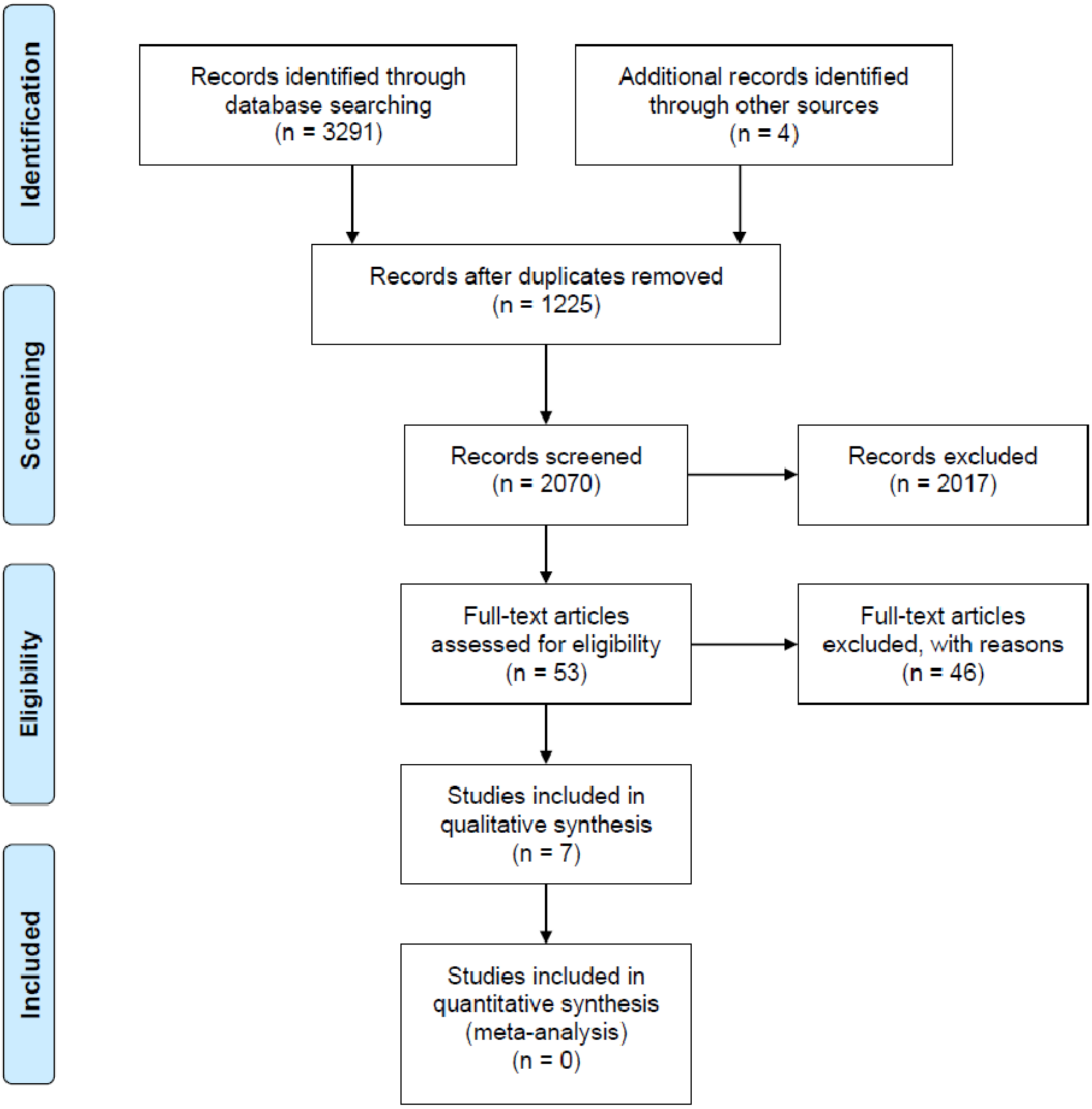
Study characteristics were clinically heterogeneous. For example, there was no consistent dependent variable that could be aggregated. As previously mentioned, a standard method for implementing telemedicine or measuring relevant outcomes is yet to exist. Thus, it was not possible to aggregate the findings in a quantitative meta-analysis. Instead, the results were synthesized based on patterns and themes that were in line with the aim of this systematic review. Decisions made using telemedicine solutions—diagnostic decision support, receiving facility decisions, and medical directions for treatment—were examined for primary outcomes of interest. Lastly, mortality, adverse events, and technical challenges were noted. The challenges not only reflected device or network-related issues, but also the effect of telemedicine on individual workflow, clinical governance, and overall organizational support.

Results

Overview

The initial search on 7 databases yielded 3291 records, where 1225 duplicates were identified. Based on the inclusion criteria, 2017 articles were excluded and the remaining 49 articles proceeded to full-text examination. Of those, 42 studies were excluded due to the setting and/or outcomes of interest. Figure 1 represents the search results as a PRISMA flowchart.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



Study Characteristics

Table S1 in Multimedia Appendix 2 shows the study characteristics of 7 selected studies, including the details of telemedicine technologies used. There were only a small number of papers identified that examine telemedicine as an intervention in the real-world setting. This suggests that despite the increasing popularity of this tool, research has been limited to simulated scenarios or halted at pilot phases. Furthermore, the methodologies reflect the nascent nature of this research field, where no framework or systematic approach exists for (1) telemedicine implementation or (2) assessment measures to evaluate its impact on clinical outcomes.

Surprisingly, none of the studies incorporated an HUD device as a component of the telemedicine system. In most interventions, smartphones or accessory equipment (eg, Bluetooth headset) enabled voice communication to a remote expert. Three studies focused their findings on two-way communication, via telephone, as the main method to deliver expert guidance [35-37]. These studies often relayed patient vital signs and other necessary information verbally, in turn gaining medical direction. For projects hoping to expand the use of telemedicine into routine use, ambulances were equipped with built-in platforms for audio and video communication, real-time vital sign transmission systems using portable devices [35,38-40]. Details about the technology used for voice communications were not provided in Dulou et al's study [37]; the authors primarily based their findings on surveys and transcribed data from after-action reviews (see Table S1 in Multimedia Appendix 2).

Bias Assessment

The risk of bias was assessed as moderate in the majority of studies, as they were nonrandomized and prospective in design (see Table S2 in Multimedia Appendix 2). Due to the type of intervention and the pre-exposure and pretraining that may be required, confounder bias was common among the studies. Examples include severity of patient illness, knowledge of intervention and performance assessment, and skill level of the paramedic. As most studies aimed to correlate telemedicine use with positive outcomes, we cannot exclude the possibility that only the most capable EMTs, clinically and technologically, were chosen for a preliminary *feasibility* phase, limiting the study's real-world applicability and validity. Information on missing data was unclear or had moderate bias in 5 studies [35,37-39,41]. Intervention bias was difficult to assess, as there appeared to be no systematic way of implementing teleconsultation and the medical direction given by a remote expert. There were trials [35,38,40] where the outcome measure was vulnerable to subjectivity, as the knowledge of telemedicine intervention could influence the interpretation of its clinical value (see Tables S2 and S3 in Multimedia Appendix 2).

Decisions Made From Using Telemedicine Solutions

Three categories under decision support were collated by the authors: diagnostic decision support, receiving facility and disposition decisions, and medical direction for prehospital treatments.

Diagnostic Decision Support

Remote expert consultation, facilitated by patient vital signs and image transfers, was useful in making accurate prehospital diagnoses. For example, evidence of diagnostic decision support was present for the majority of emergency cases where the telemedicine system was available, of which about 23% were trauma related [39]. Diagnoses were not only made using telemedicine but were made accurately, indicated by a high agreement between prehospital diagnoses via telemedicine and final in-hospital diagnoses (see κ statistics in Table S3 in Multimedia Appendix 2) [40].

Notably, usefulness of diagnostic support did not differ significantly in relation to the method of communication used. In a study comparing methods used for medical direction, opinions of consulting physicians equally favored telephone calls (mean score of 4 on a 5-point Likert scale; n=107) and full telemedicine capability with real-time data transferring and video-enabled communication (mean score of 3 on a 5-point Likert scale; n=32) [35]. EMTs also found telephone calls sufficient to relay patient status compared to using a real-time data transfer option. These findings could be explained by the introduction of added workload when using telemedicine in comparison to the simplicity of requesting a telephone consultation; however, neither a qualitative nor a quantitative explanation was offered in the study. Nevertheless, other studies [38] have reported the value of image transmission in providing an accurate situational assessment for a case discussion with remote experts.

Receiving Facility Decisions

The clinical significance of selecting an appropriate hospital destination through remote expert support is highlighted by Kim and colleagues' [41] findings. In countries where the local ambulance service is less integrated into a larger, state or national, emergency medical system, such as Korea, the receiving hospital is often determined by patients and/or their family members. Consequently, patients with mild medical conditions are often transferred to larger hospitals, overcrowding their emergency rooms, intensive care units, and general wards [41]. A real-time telemetry system (RTS) facilitated hospital destination decisions by relaying more patient information to the hospital and understanding their resource availability. Using an RTS, the destination decisions were tailored to individual patients based on medical indications [41].

Furthermore, early communication activated team responses, treatment plan decisions, and preparation for patient arrival [40]. Similarly, Bergrath et al [39] observed that in-hospital treatments could be prearranged based on prehospital notification, which ultimately reduced in-hospital time intervals and improved patient outcomes. However, such secondary outcomes were not recorded nor observed consistently across different cohorts of emergencies, which included trauma; therefore, it is difficult to assume generalizability for these findings given that patients vary in mechanisms of injury and severity.

Medical Direction for Prehospital Treatment Decisions

Medical direction for triage, treatment, and/or surgical guidance was present in all of the selected studies. These findings, however, should be interpreted in light of missing data, potential confounders, and bias in measurement of outcomes. Decision support for prehospital treatment was especially relevant for EMTs with a limited scope of skills under a regional or organizational jurisdiction. As evidenced in Germany, telemedicine technology enabled EMTs to start intravenous medication administration prior to hospital arrival with tele-doctor presence [39].

Joint treatment decisions between an on-scene first responder and remote experts were also enabled by in-built telemedicine systems. For example, the transmission of prehospital electrocardiograms and early involvement of the remote specialist ultimately shortened the door-to-balloon time in hospital [38]. The authors emphasized that such successes were made possible by an organizational approach to telemedicine with a standardized process and protocols to follow. Bergrath and colleagues [38] further found that voice communication, vital signs, and image transmission facilitated remote experts in better assessing the situation, which was important for subsequent medical directions.

Kim et al [41] developed the RTS by which patients' vital signs are transmitted to an emergency medical information center. The mean prehospital treatment time at the scene for the telemedicine group was shorter than for the control group for all patients, irrespective of the injury severity. Importantly, the authors interpreted that telemedicine was useful for receiving guidance not only on basic life support but also with advanced life-support interventions.

In particular, when there may be an experience-acquisition gap, for example, in an austere combat casualty setting, telemedicine played an invaluable role. In 14.8% of cases, telemedicine was necessary for patient management during prolonged field care. Prolonged field care requires ongoing assessment, interpreting vital trends over time, and identifying early decompensation, all of which are infrequently encountered by many prehospital providers in metropolitan areas. In such scenarios, guidance for clinical decisions, including differential diagnoses and treatment plans, were available from remote experts for frontline medics [36]. In warzone areas—Africa, Central Europe, and Afghanistan—surgical coaching and advice regarding further patient management was also made possible by telemedicine [37]. These findings are often descriptive and lack details on specific clinical outcomes. Nevertheless, the authors highlight that in its most basic form, a simple telephone consultation is highly effective and easily adopted by first responders in the military setting.

While telemedicine appears to facilitate prehospital treatment decisions, confounders (eg, paramedic skill level and patient severity) and an absence of randomized controls in many studies may overestimate a true relationship. A recent feasibility study conducted by Yperzeele and colleagues [40] reported that over 90% of teleconsultations were clinically useful, although the study did not allow the interventions to be decided by remote experts. It is difficult to understand how the remote experts were able to influence treatment plans delivered by the EMTs. Cho et al [35] used negative controls in order to demonstrate the usability of an ambulance-embedded telemedicine solution over routine use of telephone calls for making medical decisions. When comparing a telemedicine system to telephone consultation only, the authors concluded that there were no advantages in incorporating a full telemedicine capability (eg, real-time data transfer) in the ambulances. The introduction of telemedicine significantly reduced consultations for on-scene treatment, possibly due to the friction cost of using an unfamiliar system. Importantly, the case load and patient presentations were similar before and after telemedicine implementation, suggesting that case load and patient acuity were not responsible for hindering telemedicine adoption.

Overall, the results suggest that telemedicine is useful, especially in challenging and unpredictable environments. However, the methodologies used and the study characteristics make findings inconclusive. Replication is required in larger clinical trials with a randomized control group in order to better understand the usability, adoption, and reliability of telemedicine.

Mortality, Hospital Stay, Adverse Events, and Technical Challenges

No adverse events or safety concerns were raised during or after telemedicine intervention [39,40]. Yperzeele and colleagues [40] also reported no patient complaints or refusals for telemedicine use. Kim et al's [41] study was the only one to compare mortality between the study group and controls. There was no statistically significant difference between the RTS group and the control group in mortality (P=.45) and length of stay in the emergency department (P=.82). Overall mortality was examined by DeSoucy et al [36]; however, the effects of telemedicine were not considered for statistical analyses. Cho et al [35] and Bergrath et al [38] did not report on adverse events, safety, or mortality.

Telemedicine Technology Adoption and Challenges

In terms of perceived user satisfaction with telemedicine solutions, Yperzeele et al [40] measured general acceptance of the system, which was high for both teleconsultants and nurses seeking advice. These findings were determined from high rates of system activation (75.4%) and a Likert scale intended to rate user friendliness. In another study, technology adoption was measured against overall quality and clinical value, where 10 missions using video transmission were deemed to be of major clinical value [38]. Nevertheless, photos appeared to be preferred by remote experts when compared to videos, due to inconsistent transmission quality. Similarly, no significant advantage was found between telephones or video-streaming features to enable two-way communication [35].

With regard to technical challenges observed, network failures and limited processing power for video transmission and streaming most frequently disrupted telemedicine use [35,38-40]. Notably, technical issues were more difficult to overcome without protocols and overall clinical governance in place [35]. This meant that more effort was required of frontline personnel to resolve technical issues, with a consequential increase in cognitive load and a disruption of workflow.

Discussion

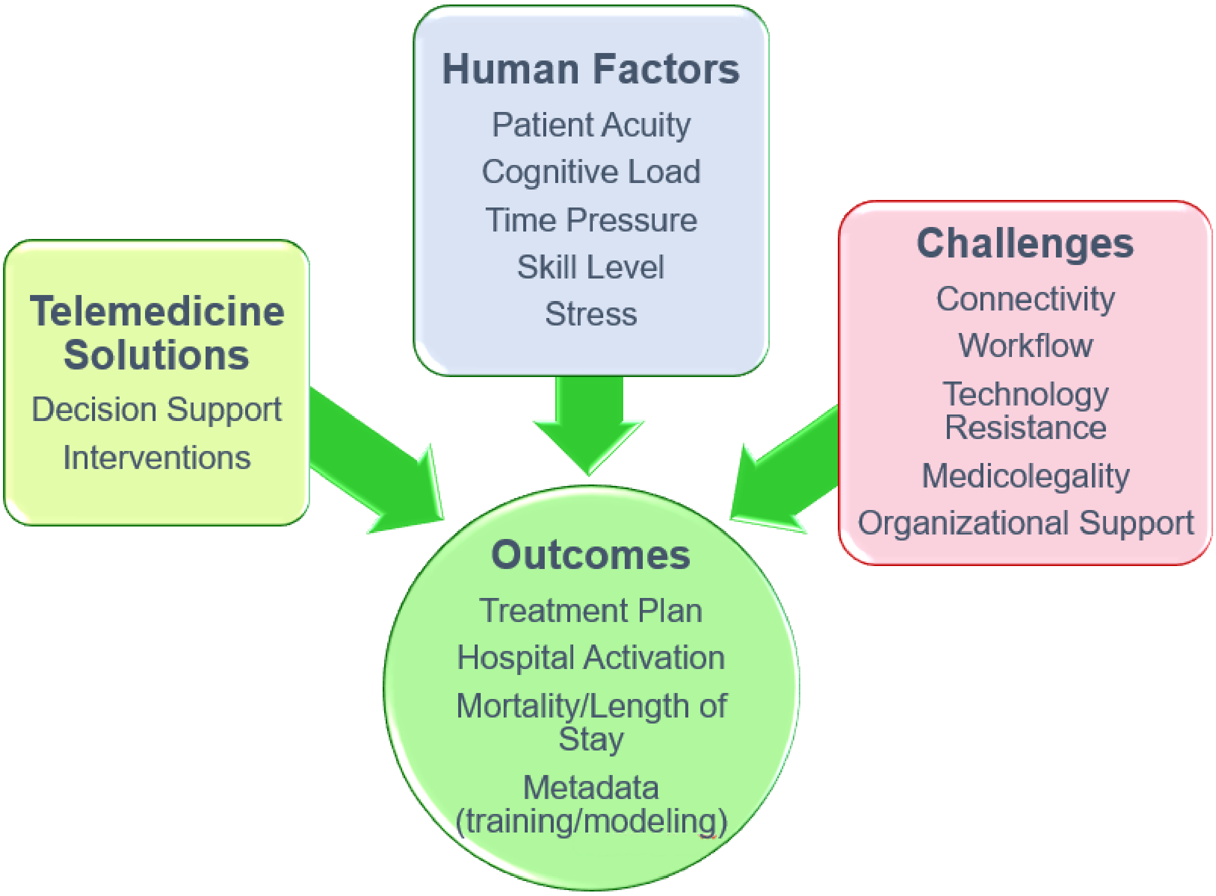
This systematic review examined whether clinical decisions are associated with telemedicine use in a prehospital emergency medical setting. To date, feasibility studies, simulated randomized controlled trials [9,26], and systematic reviews [30,32,42] on telemedicine fail to adequately address the role of decision making in patient outcomes. It is difficult to access information on (1) how telemedicine has been implemented and if any protocols exist and (2) measurements of clinical outcomes before versus after telemedicine implementation. Overall, studies investigating the effectiveness of telemedicine are heterogeneous and their findings cannot be aggregated quantitatively. The authors' interpretations of findings are typically based on observations, descriptive statistics, and surveys collected from the end users. Such methods contain inherent confounding bias, and with a small sample, often without adequate comparators, they present unavoidable challenges in delivering translatable outcomes.

Despite moderate biases present in most studies, the authors are in agreement that if employed appropriately (eg, specific patient cohort type and device and network reliability) and with organizational support, telemedicine provides invaluable access to decision support that is otherwise unobtainable. This includes support for making prehospital diagnoses [38-40], support for receiving facility decisions [38,39,41], and medical procedure guidance [35,38,41]. Medical directions appear to have the most impact in settings where EMTs require a doctor's input for medication administration and/or procedures [39,41]. The enhanced communication between EMTs and hospital medical staff, augmented by data and image transmission, all substantiate the use of telemedicine to improve the quality of continuing care.

Importantly, telemedicine may be most useful when there is an experience-acquisition gap, as there will inevitably be complex presentations or unanticipated complications. Furthermore, in rural emergency settings, the duration of patient retrieval and initial patient care can be longer [36,41]. Such scenarios may benefit from teleconsultation for ongoing monitoring and feedback on treatment plans based on real-time vital sign updates. The association between prehospital vital signs, injury severity, and in-hospital clinical outcomes highlight the importance of making prehospital data available for the severely injured [43-46]. In reality, a few vital sign data points are usually transmitted via telephone before, during, or after ambulance transfer, and whether this is adequate or not is yet to be determined.

Unfortunately, the evidence for potential downstream effects is anecdotal and not yet described in existing literature. Authors [35,36,41] identified by this systematic review anticipate downstream effects, discussed below in this section, with larger trials and further research and development on telemedicine implementation. Preliminarily, it appears that the vital sign trends received by the hospital have implications for increased readiness in patient reception, with appropriate equipment and blood products and team formation [35]. Real-time data sharing and remote expert consultation may allow for the preparation of tailored treatments and interventions prior to the patient's arrival at hospital. Shortened on-scene times may also be achieved with teleconsultation [39], with the possibility of reduction in overall transfer time [41]. Finally, improved disposition decisions were often reported, which may achieve significant cost savings through better use of emergency resources and a reduction in unnecessary secondary transfers. Figure 2 summarizes the interplay between the tasks faced by the paramedic, human factors, and telemedicine technology in influencing in-hospital and potential long-term outcomes.

Figure 2. A network of factors influencing outcomes related to hospital operation and patient care.



Based on the findings of this systematic review, future research should develop standardized clinical and technical protocols for telemedicine implementation. Bergrath and colleagues [38] highlighted the importance of an organizational approach in overcoming technical issues and achieving technology adoption. A solution for connection issues may be easier to achieve than developing clinical protocols and user case scenarios to ensure safety and systematic implementation of the technology, especially in the event of connection failure. At the very least, these findings stress the need for technical issues to be resolved, with prioritized connection to the hospital team for data and image transfer for severely injured patients. Importantly, the EMTs' workflow should be considered, such that real-time data transmission and remote expert interaction requires minimal cognitive load. To be effective, developed user case scenarios will need to target various patient presentations and paramedic experience levels and will need to account for human factors that may hinder one's decision-making processes under stress [13]. With regard to decision support, the cooperation of physicians in providing structured, measurable feedback is useful in correlating specific clinical outcomes and audits.

This systematic review has potential limitations, mostly stemming from the inclusion of nonrandomized studies. Specifically, these include difficulties in assessing, as well as having limited access to, a full data set and the risk of selective reporting in nonrandomized studies. The theoretical foundation for telemedicine solutions and the roles they play in *decision making* that are relevant to the aims of this review are also limited. In relation to the search strategy, the term *telemedicine* may have missed other decision support technological tools (eg, clinical decision support systems) with embedded remote expert communication and data-sharing capacity. Lastly, publication bias may exist for feasibility studies, as they are more likely to report positive outcomes in order to conduct subsequent clinical trials.

In conclusion, there is a role for telemedicine in supporting prehospital decision making for diagnoses, lifesaving interventions, and hospital destination. Although research in this area is in its infancy, further research into telemedicine as a tool for decision support, in real patient encounters, deserves more attention. As one of the authors highlights, "clearly, not all aspects of an emergency can be addressed by teleconsultation, *but in cases when a medical decision must be made*, it may provide a beneficial alternative" [39]. It is conceivable that telemedicine instills decision-making confidence in prehospital providers to commit to a treatment plan. An important function of the remote expert is to provide guidance when managing unanticipated events, such as during times of disaster or unprecedented virus outbreak [47,48]. In order to increase proficiency, accountability, and improved patient care, prehospital providers should consider using teleconsultations during training as well as in clinical practice.

Acknowledgments

This systematic review was funded by the Transport Accident Commission, Victoria, Australia.

Conflicts of Interest

At the time of submission of this paper, the first author (YK) was the project manager for a project titled Ambulance Victoria HUD decision-support technology usability and effectiveness evaluation.

Multimedia Appendix 1
MEDLINE search strategy.
[DOCX File , 18 KB - jmir_v22i12e18959_app1.docx]

Multimedia Appendix 2
Characteristics, risks of bias, and summary of findings of selected studies.
[DOCX File , 26 KB - jmir_v22i12e18959_app2.docx]

References

1. Fitzgerald M, Reilly S, Smit DV, Kim Y, Mathew J, Boo E, et al. The World Health Organization trauma checklist versus Trauma Team Time-out: A perspective. Emerg Med Australas 2019 Oct;31(5):882-885 [FREE Full text] [doi: 10.1111/1742-6723.13306] [Medline: 31081585]
2. Kilner E, Sheppard LA. The role of teamwork and communication in the emergency department: A systematic review. Int Emerg Nurs 2010 Jul;18(3):127-137. [doi: 10.1016/j.ienj.2009.05.006] [Medline: 20542238]
3. Valenzuela Espinoza A, De Smedt A, Guldolf K, Vandervorst F, Van Hooff R, Fernandez Tellez H, et al. Opinions and beliefs about telemedicine for emergency treatment during ambulance transportation and for chronic care at home. Interact J Med Res 2016 Mar 30;5(1):e9 [FREE Full text] [doi: 10.2196/ijmr.5015] [Medline: 27029999]
4. Joseph B, Hadeed G, Sadoun M, Rhee PM, Weinstein RS. Video consultation for trauma and emergency surgical patients. Crit Care Nurs Q 2012;35(4):341-345. [doi: 10.1097/cnq.0b013e318266c2f2]
5. Terkelsen CJ, Nørgaard BL, Lassen JF, Gerdes JC, Ankersen JP, Rømer F, et al. Telemedicine used for remote prehospital diagnosing in patients suspected of acute myocardial infarction. J Intern Med 2002 Nov;252(5):412-420 [FREE Full text] [doi: 10.1046/j.1365-2796.2002.01051.x] [Medline: 12528759]

6. Hubert GJ, Müller-Barna P, Audebert HJ. Recent advances in TeleStroke: A systematic review on applications in prehospital management and Stroke Unit treatment or TeleStroke networking in developing countries. Int J Stroke 2014 Dec;9(8):968-973. [doi: 10.1111/ijs.12394] [Medline: 25381687]
7. Mascitelli JR, Schlachter L, Chartrain AG, Oemke H, Gilligan J, Costa AB, et al. Navigation-linked heads-up display in intracranial surgery: Early experience. Oper Neurosurg (Hagerstown) 2018 Aug 01;15(2):184-193 [FREE Full text] [doi: 10.1093/ons/opx205] [Medline: 29040677]
8. Gibby JT, Swenson SA, Cvetko S, Rao R, Javan R. Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography. Int J Comput Assist Radiol Surg 2019 Mar;14(3):525-535. [doi: 10.1007/s11548-018-1814-7] [Medline: 29934792]
9. Rörtgen D, Bergrath S, Rossaint R, Beckers SK, Fischermann H, Na I, et al. Comparison of physician staffed emergency teams with paramedic teams assisted by telemedicine--A randomized, controlled simulation study. Resuscitation 2013 Jan;84(1):85-92. [doi: 10.1016/j.resuscitation.2012.06.012] [Medline: 22750663]
10. Skorning M, Bergrath S, Rörtgen D, Beckers SK, Brokmann JC, Gillmann B, Med-on-@ix-Working Group. Teleconsultation in pre-hospital emergency medical services: Real-time telemedical support in a prospective controlled simulation study. Resuscitation 2012 May;83(5):626-632. [doi: 10.1016/j.resuscitation.2011.10.029] [Medline: 22115932]
11. Schaer R, Melly T, Muller H, Widmer A. Using smart glasses in medical emergency situations, a qualitative pilot study. In: Proceedings of the 2016 IEEE Wireless Health (WH) Conference. 2016 Presented at: 2016 IEEE Wireless Health (WH) Conference; October 25-27, 2016; Bethesda, MD p. 1-5. [doi: 10.1109/wh.2016.7764556]
12. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth. Global Observatory for eHealth series - Volume 2. Geneva, Switzerland: World Health Organization; 2010. URL: https://apps.who.int/iris/bitstream/handle/10665/44497/9789241564144_eng.pdf?sequence=1 [accessed 2020-11-26]
13. Groombridge CJ, Kim Y, Maini A, Smit DV, Fitzgerald MC. Stress and decision-making in resuscitation: A systematic review. Resuscitation 2019 Nov;144:115-122 [FREE Full text] [doi: 10.1016/j.resuscitation.2019.09.023] [Medline: 31562904]
14. Fitzgerald M, Cameron P, Mackenzie C, Farrow N, Scicluna P, Gocentas R, et al. Trauma resuscitation errors and computer-assisted decision support. Arch Surg 2011 Feb;146(2):218-225. [doi: 10.1001/archsurg.2010.333] [Medline: 21339436]
15. Williams CN, Bratton SL, Hirshberg EL. Computerized decision support in adult and pediatric critical care. World J Crit Care Med 2013 Nov 04;2(4):21-28 [FREE Full text] [doi: 10.5492/wjccm.v2.i4.21] [Medline: 24701413]
16. Salinas J, Nguyen R, Darrah M, Kramer GA, Serio-Melvin ML, Mann EA, et al. Advanced monitoring and decision support for battlefield critical care environment. US Army Med Dep J 2011:73-81. [Medline: 21607909]
17. Wu F, Williams M, Kazanzides P, Brady K, Fackler J. A modular clinical decision support system: Clinical prototype extensible into multiple clinical settings. In: Proceedings of the 3rd International Conference on Pervasive Computing Technologies for Healthcare. 2009 Presented at: 3rd International Conference on Pervasive Computing Technologies for Healthcare; April 1-3, 2009; London, UK p. 1-4 URL: https://eudl.eu/pdf/10.4108/ICST.PERVASIVEHEALTH2009.6078 [doi: 10.4108/icst.pervasivehealth2009.6078]
18. Liberati EG, Ruggiero F, Galuppo L, Gorli M, González-Lorenzo M, Maraldi M, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017 Sep 15;12(1):113 [FREE Full text] [doi: 10.1186/s13012-017-0644-2] [Medline: 28915822]
19. Chowdhury NM, Finnegan P, Fitzgerald M, Chiu WK. The need for casualty care decision support in civilian and defense sectors. Procedia Eng 2017;188:301-308. [doi: 10.1016/j.proeng.2017.04.488]
20. Jones CMC, Cushman JT, Lerner EB, Fisher SG, Seplaki CL, Veazie PJ, et al. Prehospital trauma triage decision-making: A model of what happens between the 9-1-1 call and the hospital. Prehosp Emerg Care 2016;20(1):6-14 [FREE Full text] [doi: 10.3109/10903127.2015.1025157] [Medline: 26017368]
21. Newgard CD, Nelson MJ, Kampp M, Saha S, Zive D, Schmidt T, et al. Out-of-hospital decision making and factors influencing the regional distribution of injured patients in a trauma system. J Trauma 2011 Jun;70(6):1345-1353 [FREE Full text] [doi: 10.1097/TA.0b013e3182191a1b] [Medline: 21817971]
22. Perona M, Rahman MA, O'Meara P. Paramedic judgement, decision-making and cognitive processing: A review of the literature. Australas J Paramed 2019 Feb 04;16:1-12. [doi: 10.33151/ajp.16.586]
23. Evans R, McGovern R, Birch J, Newbury-Birch D. Which extended paramedic skills are making an impact in emergency care and can be related to the UK paramedic system? A systematic review of the literature. Emerg Med J 2014 Jul;31(7):594-603 [FREE Full text] [doi: 10.1136/emermed-2012-202129] [Medline: 23576227]
24. Stub D, Bernard S, Pellegrino V, Smith K, Walker T, Sheldrake J, et al. Refractory cardiac arrest treated with mechanical CPR, hypothermia, ECMO and early reperfusion (the CHEER trial). Resuscitation 2015 Jan;86:88-94. [doi: 10.1016/j.resuscitation.2014.09.010] [Medline: 25281189]
25. Deldar K, Bahaadinbeigy K, Tara A. Teleconsultation and clinical decision making: A systematic review. Acta Inform Med 2016 Jul 16;24(4):286-292 [FREE Full text] [doi: 10.5455/aim.2016.24.286-292] [Medline: 27708494]
26. Weymouth W, Thaut L, Olson N. Point of view telemedicine at point of care. Cureus 2018 Nov 30;10(11):e3662 [FREE Full text] [doi: 10.7759/cureus.3662] [Medline: 30755838]

27. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-269, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
28. Poryo M, Burger M, Wagenpfeil S, Ziegler B, Sauer H, Flotats-Bastardas M, et al. Assessment of inadequate use of pediatric emergency medical transport services: The Pediatric Emergency and Ambulance Critical Evaluation (PEACE) Study. Front Pediatr 2019 Oct 25;7:442. [doi: 10.3389/fped.2019.00442]
29. Shah MN, Cushman JT, Davis CO, Bazarian JJ, Auinger P, Friedman B. The epidemiology of emergency medical services use by children: An analysis of the National Hospital Ambulatory Medical Care Survey. Prehosp Emerg Care 2008;12(3):269-276 [FREE Full text] [doi: 10.1080/10903120802100167] [Medline: 18584491]
30. Amadi-Obi A, Gilligan P, Owens N, O'Donnell C. Telemedicine in pre-hospital care: A review of telemedicine applications in the pre-hospital environment. Int J Emerg Med 2014;7:29 [FREE Full text] [doi: 10.1186/s12245-014-0029-0] [Medline: 25635190]
31. Eder P, Reime B, Wurmb T, Kippnich U, Shammas L, Rashid A. Prehospital telemedical emergency management of severely injured trauma patients. Methods Inf Med 2018 Nov;57(5-06):231-242. [doi: 10.1055/s-0039-1681089] [Medline: 30875702]
32. Hasselberg M, Beer N, Blom L, Wallis LA, Laflamme L. Image-based medical expert teleconsultation in acute care of injuries. A systematic review of effects on information accuracy, diagnostic validity, clinical outcome, and user satisfaction. PLoS One 2014;9(6):e98539 [FREE Full text] [doi: 10.1371/journal.pone.0098539] [Medline: 24887257]
33. Winburn AS, Brixey JJ, Langabeer J, Champagne-Langabeer T. A systematic review of prehospital telehealth utilization. J Telemed Telecare 2017 Jun 22;24(7):473-481. [doi: 10.1177/1357633x17713140]
34. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
35. Cho SJ, Kwon IH, Jeong J. Application of telemedicine system to prehospital medical control. Healthc Inform Res 2015 Jul;21(3):196-200 [FREE Full text] [doi: 10.4258/hir.2015.21.3.196] [Medline: 26279957]
36. DeSoucy E, Shackelford S, DuBose JJ, Zweben S, Rush SC, Kotwal RS, et al. Review of 54 cases of prolonged field care. J Spec Oper Med 2017;17(1):121-129. [Medline: 28285490]
37. Dulou R, Dagain A, Delmas J, Lambert E, Blondet E, Goasguen O, et al. The French mobile neurosurgical unit. Neurosurg Focus 2010 May;28(5):E13. [doi: 10.3171/2010.2.focus1016]
38. Bergrath S, Rörtgen D, Rossaint R, Beckers SK, Fischermann H, Brokmann JC, et al. Technical and organisational feasibility of a multifunctional telemedicine system in an emergency medical service - An observational study. J Telemed Telecare 2011;17(7):371-377. [doi: 10.1258/jtt.2011.110203] [Medline: 21933897]
39. Bergrath S, Czaplik M, Rossaint R, Hirsch F, Beckers SK, Valentin B, et al. Implementation phase of a multicentre prehospital telemedicine system to support paramedics: Feasibility and possible limitations. Scand J Trauma Resusc Emerg Med 2013 Jul 11;21:54 [FREE Full text] [doi: 10.1186/1757-7241-21-54] [Medline: 23844941]
40. Yperzeele L, Van Hooff R, De Smedt A, Valenzuela Espinoza A, Van Dyck R, Van de Casseye R, et al. Feasibility of AmbulanCe-Based Telemedicine (FACT) study: Safety, feasibility and reliability of third generation in-ambulance telemedicine. PLoS One 2014;9(10):e110043 [FREE Full text] [doi: 10.1371/journal.pone.0110043] [Medline: 25343246]
41. Kim YK, Kim KY, Lee KH, Kim SC, Kim H, Hwang SO, et al. Clinical outcomes on real-time telemetry system in developing emergency medical service system. Telemed J E Health 2011 May;17(4):247-253. [doi: 10.1089/tmj.2010.0152] [Medline: 21480786]
42. Eder P, Reime B, Wurmb T, Kippnich U, Shammas L, Rashid A. Prehospital telemedical emergency management of severely injured trauma patients. Methods Inf Med 2018 Nov;57(5-06):231-242. [doi: 10.1055/s-0039-1681089] [Medline: 30875702]
43. Nakada T, Masunaga N, Nakao S, Narita M, Fuse T, Watanabe H, et al. Development of a prehospital vital signs chart sharing system. Am J Emerg Med 2016 Jan;34(1):88-92. [doi: 10.1016/j.ajem.2015.09.048] [Medline: 26508581]
44. Sartorius D, Le Manach Y, David J, Rancurel E, Smail N, Thicoïpé M, et al. Mechanism, Glasgow Coma Scale, Age, and Arterial Pressure (MGAP): A new simple prehospital triage score to predict mortality in trauma patients. Crit Care Med 2010 Mar;38(3):831-837. [doi: 10.1097/CCM.0b013e3181cc4a67] [Medline: 20068467]
45. Raux M, Sartorius D, Le Manach Y, David J, Riou B, Vivien B. What do prehospital trauma scores predict besides mortality? J Trauma 2011;71(3):754-759. [doi: 10.1097/ta.0b013e3181fd0dae]
46. Bruijns SR, Guly HR, Wallis LA. Effect of spinal immobilization on heart rate, blood pressure and respiratory rate. Prehosp Disaster Med 2013 Jun;28(3):210-214. [doi: 10.1017/S1049023X13000034] [Medline: 23522699]
47. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine during the COVID-19 pandemic: Experiences from Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
48. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: A call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]

Abbreviations

EMT:	emergency medical technician
HUD:	heads-up display
ICT:	information and communication technology
PRISMA:	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROBINS-I:	Risk Of Bias In Non-randomised Studies–of Interventions
RTS:	real-time telemetry system

Edited by G Eysenbach; submitted 29.03.20; peer-reviewed by H Inaba, I DeSouza; comments to author 12.06.20; revised version received 16.07.20; accepted 28.10.20; published 08.12.20.

Please cite as:
Kim Y, Groombridge C, Romero L, Clare S, Fitzgerald MC
Decision Support Capabilities of Telemedicine in Emergency Prehospital Care: Systematic Review
J Med Internet Res 2020;22(12):e18959
URL: http://www.jmir.org/2020/12/e18959/
doi:10.2196/18959
PMID:33289672

Review

Smartphone Overuse and Visual Impairment in Children and Young Adults: Systematic Review and Meta-Analysis

Jian Wang[1], MSc; Mei Li[2], MD, PhD; Daqiao Zhu[1], PhD; Yang Cao[3], PhD

[1]School of Nursing, Shanghai Jiao Tong University, Shanghai, China
[2]School of Medical Sciences, Örebro University, Örebro, Sweden
[3]Clinical Epidemiology and Biostatistics, School of Medical Sciences, Örebro University, Örebro, Sweden

Corresponding Author:
Yang Cao, PhD
Clinical Epidemiology and Biostatistics
School of Medical Sciences
Örebro University
Södra Grev Rosengatan 1
Örebro, 70182
Sweden
Phone: 46 196026236
Email: yang.cao@oru.se

Abstract

Background: Smartphone overuse has been cited as a potentially modifiable risk factor that can result in visual impairment. However, reported associations between smartphone overuse and visual impairment have been inconsistent.

Objective: The aim of this systematic review was to determine the association between smartphone overuse and visual impairment, including myopia, blurred vision, and poor vision, in children and young adults.

Methods: We conducted a systematic search in the Cochrane Library, PubMed, EMBASE, Web of Science Core Collection, and ScienceDirect databases since the beginning of the databases up to June 2020. Fourteen eligible studies (10 cross-sectional studies and 4 controlled trials) were identified, which included a total of 27,110 subjects with a mean age ranging from 9.5 to 26.0 years. We used a random-effects model for meta-analysis of the 10 cross-sectional studies (26,962 subjects) and a fixed-effects model for meta-analysis of the 4 controlled trials (148 subjects) to combine odds ratios (ORs) and effect sizes (ES). The I^2 statistic was used to assess heterogeneity.

Results: A pooled OR of 1.05 (95% CI 0.98-1.13, *P*=.16) was obtained from the cross-sectional studies, suggesting that smartphone overuse is not significantly associated with myopia, poor vision, or blurred vision; however, these visual impairments together were more apparent in children (OR 1.06, 95% CI 0.99-1.14, *P*=.09) than in young adults (OR 0.91, 95% CI 0.57-1.46,*P*=.71). For the 4 controlled trials, the smartphone overuse groups showed worse visual function scores compared with the reduced-use groups. The pooled ES was 0.76 (95% CI 0.53-0.99), which was statistically significant (*P*<.001).

Conclusions: Longer smartphone use may increase the likelihood of ocular symptoms, including myopia, asthenopia, and ocular surface disease, especially in children. Thus, regulating use time and restricting the prolonged use of smartphones may prevent ocular and visual symptoms. Further research on the patterns of use, with longer follow up on the longitudinal associations, will help to inform detailed guidelines and recommendations for smartphone use in children and young adults.

(J Med Internet Res 2020;22(12):e21923) doi:10.2196/21923

KEYWORDS

visual impairment; smartphone; mobile phone; overuse; child; young adult; systematic review; meta-analysis

Introduction

The use of smartphones has been increasing rapidly since their introduction in the late 2000s [1]. In 2019, the global smartphone penetration had reached approximately 41.5% of the global population [2]. Notably, the number of smartphone users in China was around 700 million in 2018, accounting for half of the Chinese population [3]. In addition, more than 80% of people in the United Kingdom owned or had ready access to a smartphone in 2019, representing a significant increase from

50% in 2012 [4]. Furthermore, more than 90% of young people between 16 and 34 years old in the United Kingdom owned a smartphone in 2019 [4].

With the continuous rise in youth digital media consumption, the incidence of ocular problems has also dramatically increased. A large portion of the population currently suffers from visual impairment, especially in Asian countries, with a rapidly increasing prevalence and younger age of onset [5-8]. It has been estimated that 49.8% (4.8 billion) and 9.8% (0.9 billion) of the global population will have myopia or high myopia by 2050 [9]. A recent study indicated that about 60 years ago, only 10%-20% of the Chinese population was nearsighted, but the percentage reached up to 90% of teenagers and young adults in 2015 [10]. Consistently, a school-based retrospective longitudinal cohort study (N=37,424 participants) found that the prevalence of myopia significantly increased from 56% in 2005 to 65% in 2015 [8].

Therefore, smartphone overuse among children and young adults has become a matter of crucial concern [11-13]. Several studies found increased use of digital devices in children aged 2-11 years old [14,15]. For example, a study including children aged 9-11 years from 12 countries showed that 54.2% of the children exceeded proposed screen time guidelines (≤2 hours per day) [15]. Compared with older people, children and young adults have greater risks of the undesirable consequences of smartphone overuse because they have less self-control in smartphone use [11]. A cross-sectional study (N=2639 participants) indicated that 22.8% of teenagers were addicted to smartphone use, which was related to hypertension [16]. Another study showed that users of mobile devices spent >20 hours weekly on email, text messages, and social networking services, indicating the heavy reliance on smartphones in their communication with other people [17]. Overall, smartphone overuse may result in significant harmful physical, psychological, and social consequences [18,19].

Some experimental studies have indicated that long-term use of a smartphone plays a key role in visual impairment, increasing the likelihood of poor vision [20-22]. For instance, a prospective clinical study (N=50 participants) showed that smartphone use for 4 hours resulted in a higher ocular surface disease index than that measured at baseline [20]. Kim et al [23] found that the increase of ocular symptoms extended to the general population, especially in adolescents, after expansion of smartphone use. However, other studies have reported the lack of evidence for such an association [24]. For example, a cross-sectional study (N=1153 participants) using stratified random cluster samples did not find a statistically significant association between smartphone use time and myopia [25]. Similarly, a study conducted in Ireland (N=418 participants) indicated that smartphone use time was not a risk factor for myopia [26]. Toh et al [27] found that smartphone use time was associated with an increased risk of visual symptoms (ie, blurring of vision, dry eye), but a decreased odds of myopia.

Despite increased concern about impaired vision due to smartphone overuse, existing quantitative evidence about the relationships between excessive smartphone use and visual impairment remains equivocal. Therefore, it is essential to confirm and quantify whether excessive smartphone use may result in visual impairment, especially in children and young adults.

The aim of this study was to conduct a systematic review and meta-analysis to summarize the existing evidence on the associations between smartphone overuse and visual impairment in children and young adults, which may further guide potential interventions to reduce the harmful impact of smartphone overuse on vision in this susceptible subpopulation.

Methods

Data Sources and Search Strategy

This systematic review and meta-analysis was based on a protocol designed in line with the standard Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [28] and Meta-analysis of Observational Studies in Epidemiology (MOOSE) [29] criteria.

A systematic search was carried out in PubMed (US National Library of Medicine), Embase (Wolters Kluwer Ovid), Web of Science Core Collection (Clarivate Analytics), ScienceDirect (Elsevier), and Cochrane library (John Wiley & Sons, Ltd) for observational and experimental studies that investigated smartphone overuse or addiction in children (aged<18 years) or young people (aged<40 years), and its associations with impaired visual function such as myopia, poor vision, or blurred vision. To minimize publication bias, we also searched for additional studies in grey literature sources, including Virtual Health Library [30], NARCIS [31], Grey literature report [32], and Open grey EU [33]. The search was limited to publications published in English.

Free text and Medical Subject Headings (MeSH) terms were used for the search, including phone, smartphone, mobile/cell/cellular phone, electronic device, use, use time, screen time, overuse, addiction, eye, visual acuity, vision, vision screening, eyesight, myopia, myopic refraction, shortsighted/nearsighted/short sight, near sight, refraction errors, ocular/health effect, optic, blind, ophthalmology, optometry, retina, ametropia/amblyopia symptom, visual assessment, and visual problem (see Multimedia Appendix 1 for the complete search strategy). We included all observational studies and controlled trials (randomized or nonrandomized) addressing smartphone use and visual impairment in humans since the beginning of the databases up to June 2020. Furthermore, manual retrieval was performed following the initial database search to ensure the inclusion of the latest literature.

Inclusion and Exclusion Criteria

All observational and experimental studies were included if they fulfilled the following criteria: (1) original studies examining the use of a smartphone (or mobile phone) and eyesight, including population-based longitudinal studies, cohort studies, case-control studies, cross-sectional studies, and controlled clinical trials; (2) participants are children aged ≤18 years or young people aged ≤40 years (a young adult was defined as the developmental stage between 18 and 40 years [34,35]); (3) reported frequency or time of smartphone use (in minutes or hours, or per day or week); (4) the endpoint of interest is the incidence of visual impairment or decline, including myopia, poor vision, blurred vision, various visual function scores indicating impaired vision, or other unspecific visual impairments; and (5) vision measurements of the groups are provided to calculate the effect size (ES) of visual impairment or odds ratio (OR) for the risk of visual impairment, as well as the associated 95% CIs or other data to estimate the variance or accuracy (eg, standard error).

Studies were excluded if they: (1) were narrative reviews, editorial papers, commentaries, letters, or methodological papers; (2) evaluated visual function with no reliable/relevant estimates for smartphone use; (3) no reference or control group was included in the analysis; and (4) animal studies.

Data Extraction

After the systematic search of the relevant articles in the databases, two investigators (JW and ML) embarked on screening and identification of potentially relevant abstracts independently. For any disagreements that occurred between the two investigators regarding the eligibility of a study, there was a thorough discussion or advice from an academic expert (YC). Subsequently, articles for selected abstracts were downloaded, and data were extracted by JW and YC independently using a standardized form in Microsoft Excel. The extracted data were compared and summarized to obtain one final document from which the analysis was conducted. The information extracted included: name of first author, year of publication, study design, duration of study, country that the study was conducted in, eyesight measurement, smartphone use time, smartphone use frequency, sample size, incidence of cases with impaired vision, outcome ascertainment method, OR or ES and the associated 95% CI, and statistical analysis method used.

Study Quality Assessment

The Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Analytical Cross Sectional Studies, JBI Appraisal Checklist for Quasi-Experimental Studies, and JBI Critical Appraisal Checklist for Randomized Controlled Trials were used to assess the quality of the studies included in the meta-analysis [36]. JW and YC assessed the quality of the articles independently and the final assessment was achieved upon discussion (Multimedia Appendix 2).

Statistical Analysis

For studies that did not report the OR, it was calculated using the numbers of cases with and without visual impairment of the reference/control group and overuse group. For studies that measured visual impairment using continuous variables, ES was calculated as the difference between the means divided by the pooled SD as follows [37]:

where n_1 and n_2, and S_1 and S_2 are the sample sizes and standard deviations for group 1 and group 2, respectively.

A positive ES indicates a worse visual function. Heterogeneity of the included studies was investigated using the I^2 statistic [38], in which $I^2>30\%$ was considered to indicate moderate heterogeneity and $I^2>50\%$ was considered to indicate substantial heterogeneity [39]. A *P* value <.05 from the noncentral chi-squared test for heterogeneity was considered to indicate statistically significant heterogeneity [40]. The contribution of each study to the heterogeneity and their influence on the pooled OR or ES were assessed using the Baujat plot [41]. The pooled ORs with corresponding 95% CIs were calculated using random-effects models owing to heterogeneity among the studies and are presented using forest plots [42]. The possibility of publication bias was assessed by the combination of the Egger test and visual inspection of the funnel plot [43].

Subgroup analysis was performed for the cross-sectional studies according to the outcome of visual impairment (myopia, poor vision, or blurred vision) and mean age of the subjects (children, ≤18 years; young people, 18-40 years). Leave-one-out (LOO) analysis was also performed to investigate the influence of a single study on the pooled effect as an additional sensitivity analysis [44].

A two-sided *P* value <.05 of the pooled estimates was considered statistically significant unless otherwise specified. All analyses were performed in R 4.0.0 (R Foundation for Statistical Computing, Vienna, Austria) using the packages meta 4.12-0 [45] and dmetar 0.0.9000 [46].

Results

Characteristics of Included Studies

In total, 1961 articles were obtained from all of the databases. After removing duplicates, 1796 articles remained, 121 of which were considered to be relevant for the meta-analysis after screening of titles and abstracts. After screening the full text of the 121 articles downloaded, 14 articles met our inclusion criteria, including 10 cross-sectional studies and 4 controlled trials, comprising a total of 27,110 participants with mean ages ranging from 9.5 to 26.0 years. The flowchart of article searching and screening is shown in Figure 1. The 10 cross-sectional studies addressed incidents of myopia [24-27,47], blurred vision [48-50], and poor vision and other unspecified visual impairments [23,27,48,51]. In our analysis, the unspecified visual impairments were treated as poor vision. There were 2 studies [27,48] that addressed two visual impairment outcomes, and each outcome was treated as a single study in the meta-analysis. The 4 studies that used a controlled trial design assessed the ocular surface disease index score [20], asthenopia score [21], oculomotor function [52], and viewing distance [22]. A more detailed summary of the characteristics of the included studies and the main outcomes is provided in Table 1 and Table 2, respectively.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for screening and selection of articles on smartphone overuse and visual impairment in children and young adults.

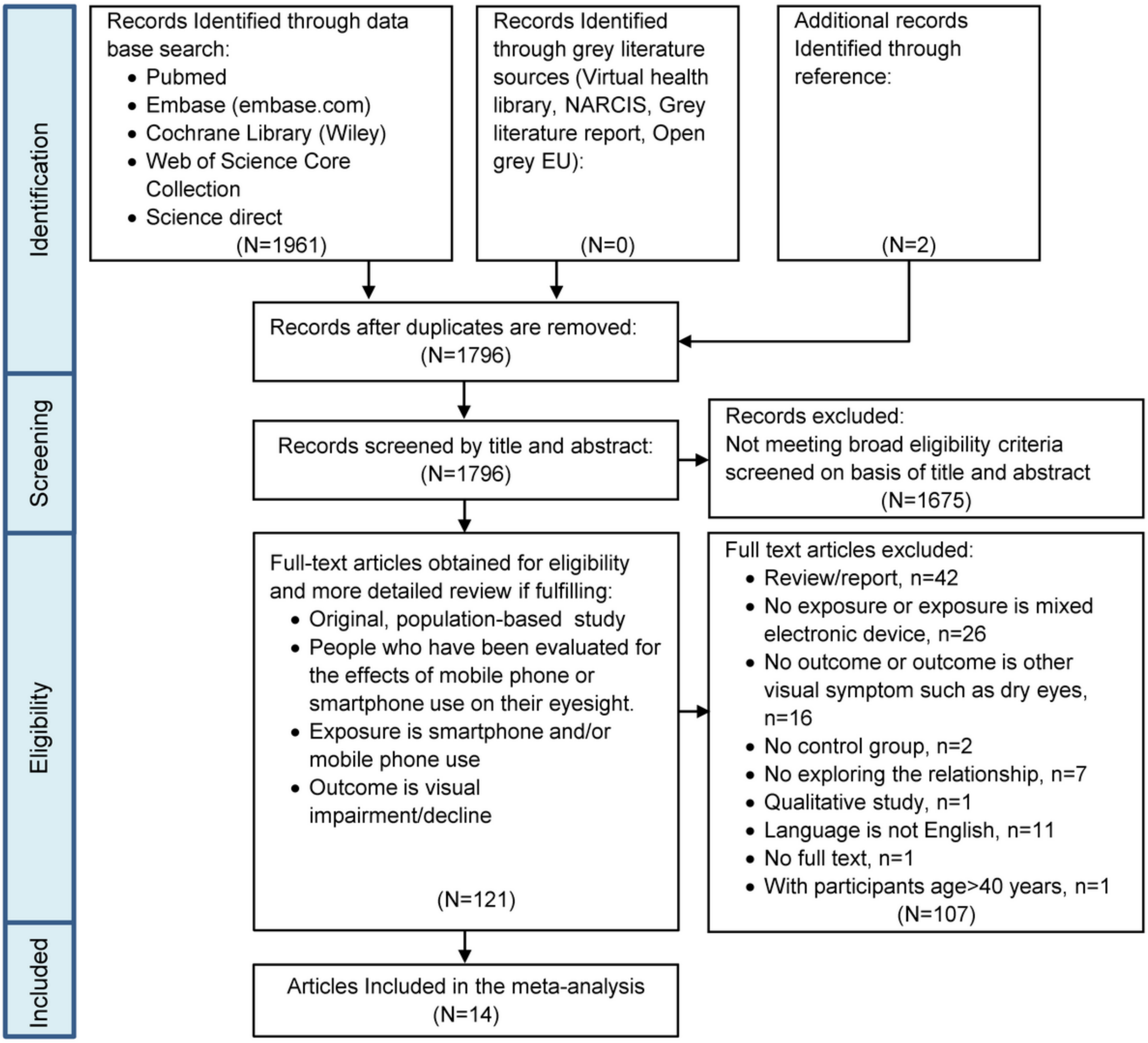


Table 1. General characteristics of the included studies.

Reference	Year	Country	Study design	Age of participants (years), mean (SD) or range	Sampling method	N participants
Küçer et al [49]	2008	Turkey	Cross-sectional	University students (age not given)	Convenience sample	229
Toh et al [27]	2019	Singapore	Cross-sectional	13.3 (2.0)	Matrix-stratified sample	1884
Merrie et al [51]	2019	Ethiopia	Cross-sectional	13.1 (2.8)	Multistage sampling	601
Guan et al [47]	2019	China	Cross-sectional	10.6 (1.15)	Randomly selected sample	19,934
Kim et al [23]	2016	Korea	Cross-sectional	15 (0.9)	Convenience sample	715
Liu et al [24]	2019	China	Cross-sectional	9.5 (2.1)	Stratified cluster sample	566
Meo et al [50]	2005	Saudi Arabia	Cross-sectional	26.0 (13.4)	Voluntary (response) sample	873
Alharbi et al [48]	2019	Saudi Arabia	Cross-sectional	21.8 (2.4)	Random sample	605
Huang et al [25]	2019	China	Cross-sectional	19.6 (0.9)	Stratified random cluster sample	1153
McCrann et al [26]	2020	Ireland	Cross-sectional	16.8 (4.4)	Voluntary sample	402
Antona et al [21]	2018	Spain	RCT[a]	23.7 (2.6)	Random sample	54
Choi et al [20]	2018	South Korea	CT[b]	26.0 (3.0)	Nonrandomized sample	50
Lee et al [52]	2019	Korea	CT	20-29	Voluntary sample	26
Long et al [22]	2017	Australia	CT	21.5 (3.3)	Voluntary sample	18

[a] RCT: randomized controlled trial.

[b] CT: controlled trial.

Table 2. Outcomes and results of the included studies.

Reference	Response rate	Exposure; type of measure	Outcome; type of measure	Main results
Küçer et al [49]	100%	Time of mobile phone possession; Q[a]	Blurred vision; Q	≤2 years: 8.8% (4/45) >2 years: 27.2% (50/184)
Toh et al [27]	93.78% (1884/2009)	Time of smartphone use (per hour); Q	(1) Myopia; Q (2) Poor vision/visual impairment; Q	(1) OR[b] 0.97 (95% CI 0.94-0.99) (2) OR 1.05 (95% CI 1.02-1.08)
Merrie et al [51]	95.09% (601/632)	Duration of mobile exposure; Q	Poor vision/visual impairment; objective assessment	>2 h/day: 6.6% (18/271) ≤ 2 h/day: 7.5% (20/265)
Guan et al [47]	UK[c]	Time of smartphone use; Q	Visual acuity; objective assessment	1 h/day: 20% (117/584); ≤1 h/day: 18% (3492/19350)
Kim et al [23]	97.41% (715/734)	Time of smartphone use; Q	Poor vision/ocular symptom score; Q	>2 h/day: 72% (260/360); ≤2 h/day: 52% (170/327)
Liu et al [24]	88.7% (566/638)	Time of smartphone use (per hour); Q	Myopia; objective assessment	OR 0.90 (95% CI 0.57-1.43)
Meo et al [50]	100%	Use of mobile phone (duration of calls); Q	Blurred vision; Q	>0.5 h/day: 5% (5/100); ≤0.5 h/day: 5.23% (39/746)
Alharbi et al [48]	93.1% (605/650)	Duration of Smartphone use per day; Q	(1) Poor vision; Q (2) Blurred vision; Q	(1) >3 h/day: 57.2% (270/472); ≤3 h/day: 45.9% (61/133） (2) >3 h/day: 46.0% (217/472); ≤3 h/day: 57.1% (76/133）
Huang et al [25]	96.08% (1153/1200)	Duration of daily smartphone use; Q	Myopia; objective assessment	>3 h/day: 84.57% (296/350); ≤ 3 h/day: 88.03% (537/610)
McCrann et al [26]	96.17% (402/418)	Time on phone (minutes/day); Q	Myopia; Q	OR 1.026 (95% CI 1.001-1.051)
Antona et al [21]	100%	Smartphone reading vs printed hardcopy reading	Asthenopia score; Q	27.96 (SD 20.11) vs 13.25 (SD 12.76)
Choi et al [20]	100%	Smartphone use after 4 hours vs baseline	Ocular surface disease index scores; Q	25.03 (SD 10.61) vs 15.08 (SD 8.83)
Lee et al [52]	86.67% (26/30)	Smartphone use 20 minutes vs 5 minutes	Oculomotor function; Q	6.35 (SD 3.54) vs 3.73 (SD 4.09)
Long et al [22]	100%	Using smartphone after 1 hour vs baseline	Viewing distance; objective assessment	27.8 (SD 7.7) cm vs 31 (SD 8.2) cm

[a] Q: questionnaire.

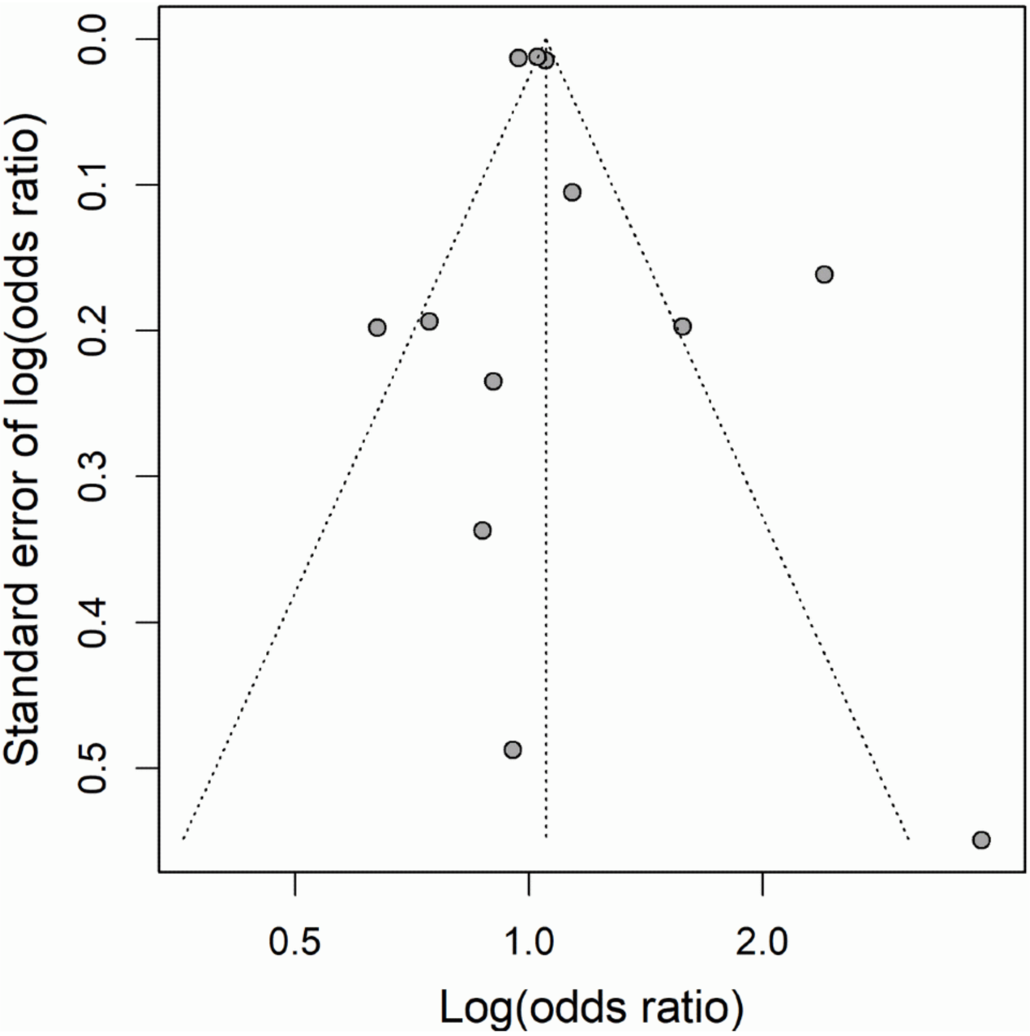
[b] OR: odds ratio.

[c] UK: unknown.

Association Between Smartphone Overuse and Incidence of Visual Impairment

The funnel plot of ORs for the included cross-sectional studies appeared to be symmetric (Figure 2). Although ORs from two studies [23,49] showed slight bias with other studies, no statistically significant publication bias was found based on the Egger test (*P*=.43).

Figure 2. Funnel plot with pseudo 95% confidence limit for cross-sectional studies.



Statistically significant heterogeneity was present among the ORs on visual impairment incidence (I^2=84%, *P*<.001; Figure 3). The Baujat plot indicated that the study of Kim et al [23] substantially contributed to the heterogeneity but had a minimal influence on the pooled OR (Figure 4). Overall, although the pooled OR showed that the odds of visual impairment was higher for the smartphone overuse group compared to the reduced-use group (OR 1.05, 95% CI 0.98-1.13), the result was not statistically significant (*P*=.16; Figure 3). None of the pooled ORs for specific visual impairment was significant in subgroup analyses. The pooled ORs for myopia, poor vision, and blurred vision were 1.00 (95% CI 0.95-1.05), 1.40 (95% CI 0.87-2.23), and 1.21 (95% CI 0.44-3.28), respectively (Figure 3). The pooled OR was not statistically significant in either of the age subgroups, which was 1.06 (95% CI 0.99-1.14, *P*=.09) for children and 0.91 (95% CI 0.57-1.46, *P*=.71) for young adults.

Figure 3. Pooled odds ratios (ORs) of visual impairment in the smartphone overuse group compared to the reduced-use group.

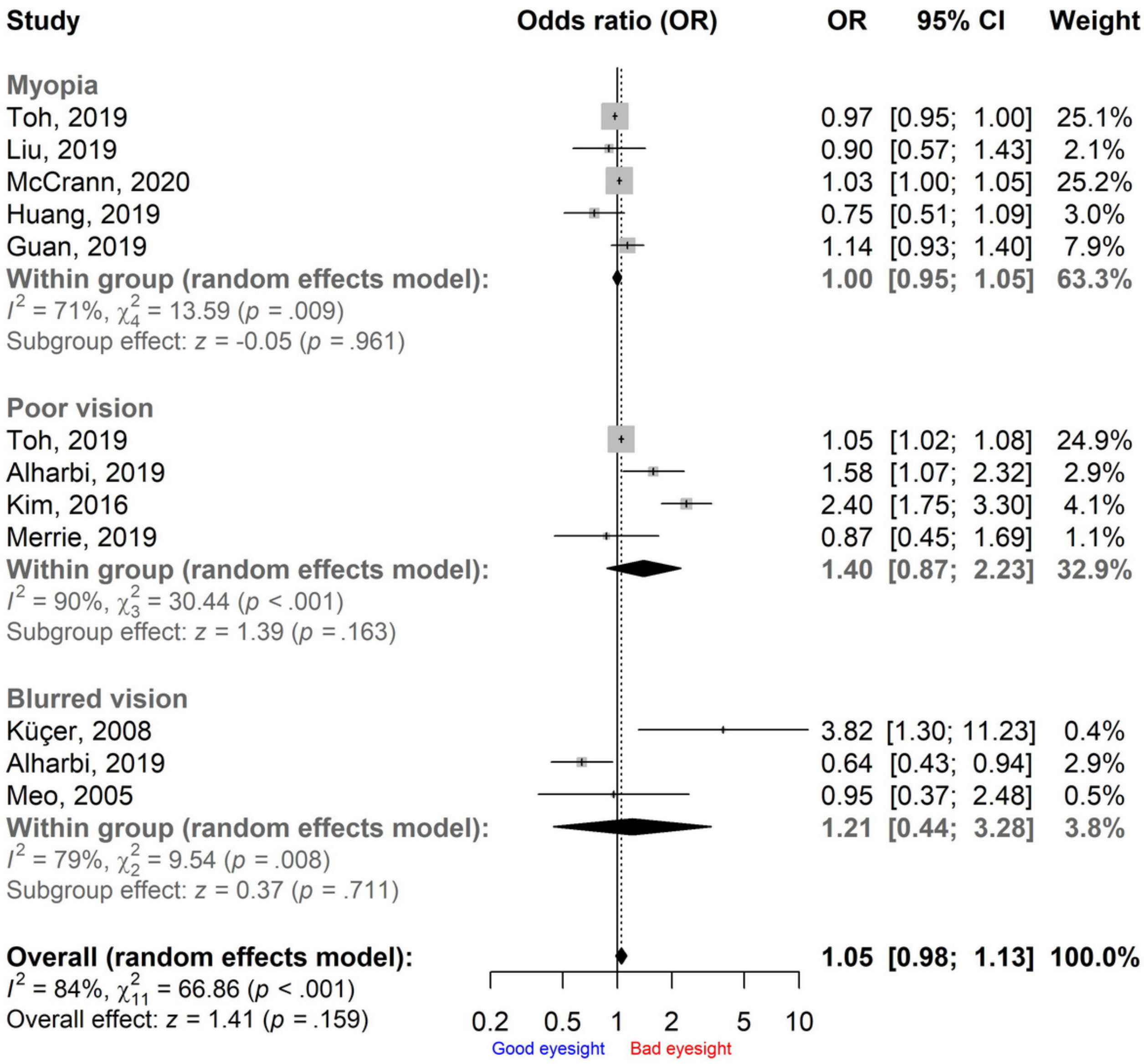
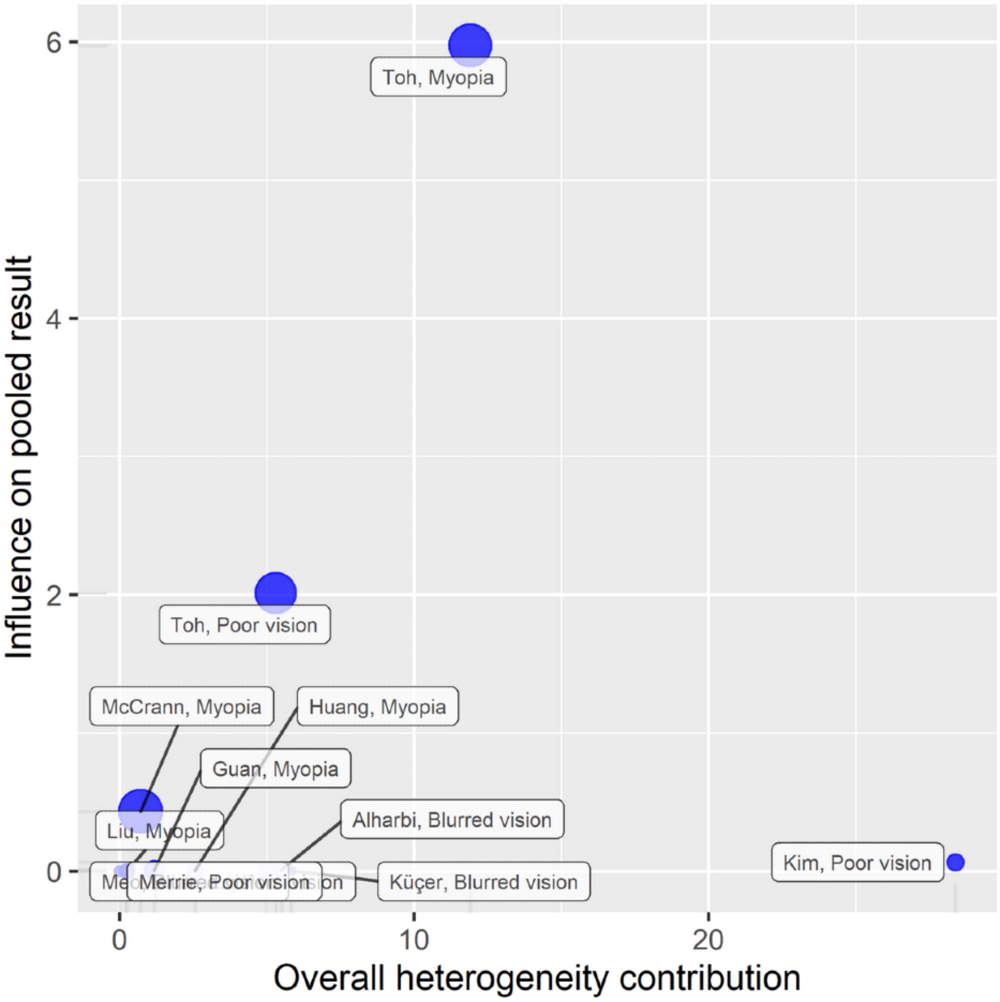
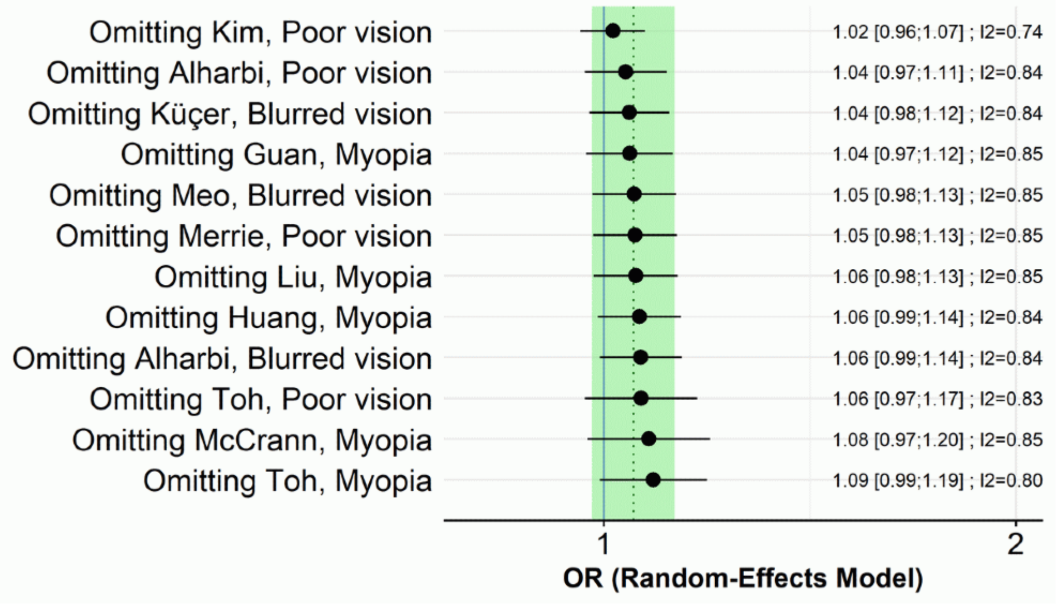


Figure 4. Baujat plot for cross-sectional studies.



The LOO sensitivity test indicated that ORs of visual impairment in the smartphone overuse group compared to the reduced-use group ranged from 1.02 to 1.09; however, none of the ORs was statistically significant (Figure 5).

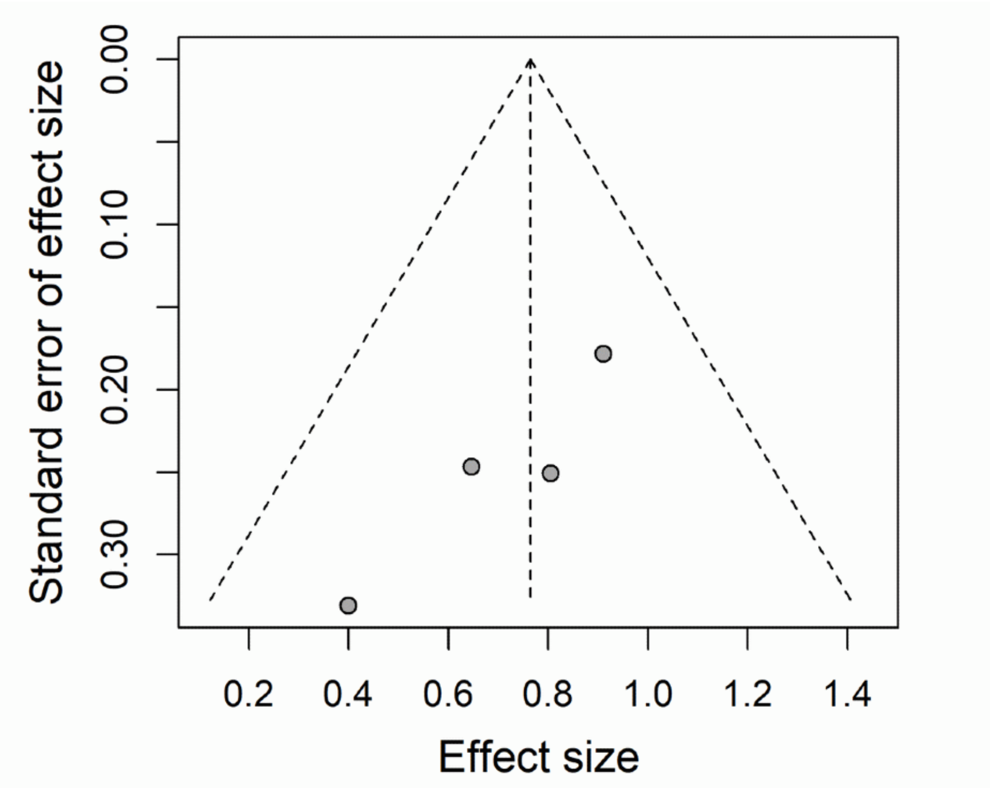
Figure 5. Pooled odds ratios (ORs) of visual impairment in the smartphone overuse group compared to the reduced-use group from leave-one-out analysis.



Smartphone Overuse Associated With Worse Visual Function Scores

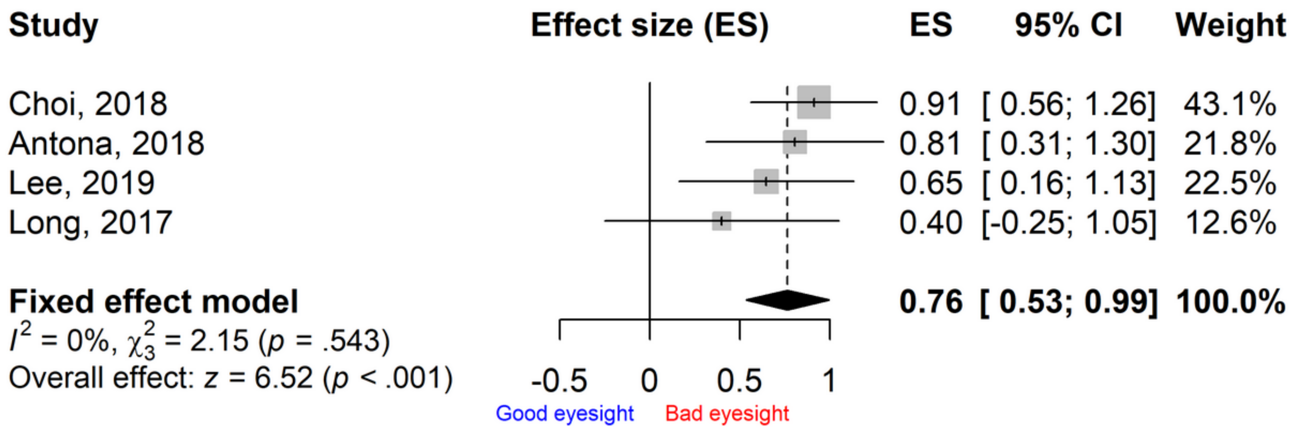
The funnel plot of ES for the included controlled trials appeared to be symmetric (Figure 6), and no statistically significant publication bias was found by the Egger test (P=.067). No statistically significant heterogeneity was present among the ESs on visual impairment incidence (I^2=0%, P=.54; Figure 7).

Figure 6. Funnel plot with pseudo 95% confidence limit for controlled trials.



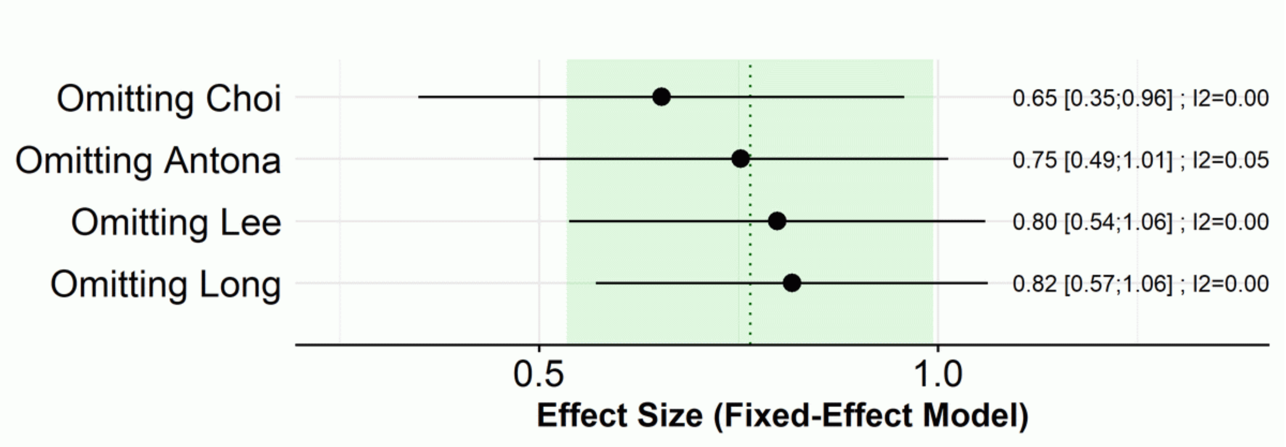
In all of the controlled trials, the smartphone overuse group showed worse visual function scores than the reduced-use group, with ESs ranging from 0.40 to 0.91 (Figure 7). The pooled ES was 0.76 (95% CI 0.53-0.99), which was statistically significant (P<.001), indicating that compared with the reduced-use group, the visual function score in the smartphone overuse group was 0.76 SD worse (Figure 7).

Figure 7. Pooled effect size (ES) of visual function score in the smartphone overuse group compared to the reduced-use group.



The LOO sensitivity test indicated that the results are robust, with the ESs ranging from 0.65 to 0.82, and all of the ESs were statistically significant (Figure 8).

Figure 8. Pooled effect sizes (ESs) of visual function score in the smartphone overuse group compared to the reduced-use group from leave-one-out analysis.



Discussion

Principal Findings

The purpose of this systematic review and meta-analysis was to summarize currently available evidence with reference to the relationship between smartphone overuse and visual impairment in children and young adults. Among the 14 studies included in the analysis, 9 found a significant association between smartphone overuse and visual impairment. Our pooled results showed negative but not statistically significant associations (OR=1.05, 95% CI 0.98-1.13) between smartphone overuse and myopia, blurred vision, or poor vision in the included cross-sectional studies. However, the adverse effect was more apparent in children (OR=1.06, 95% CI 0.99-1.14) than in young adults (OR=0.91, 95% CI 0.57-1.46). We also found that smartphone overuse may cause worse visual function than reduced use in the included controlled trials (ES=0.76, 95% CI 0.53-0.99). As the results are mixed, further studies are warranted. To our knowledge, this is the first systematic review that comprehensively summarized existing data on smartphone overuse and visual impairment in children and young adults.

There are several possible reasons for the lack of a statistically significant association observed between smartphone overuse and visual impairment when pooling cross-sectional studies. First, most of the existing studies included in this systematic review were from Asia, which has higher prevalence rates of visual impairment. The myopia prevalence in East Asia was already reported to be high before the introduction of digital devices [53]. Previous studies indicated that myopia prevalence increased more rapidly in people with more years of education and intensive schooling without particular exposure to screen devices [54-56]. For example, a study conducted in Singapore found that myopia prevalence increased more rapidly among individuals who started elementary school after the 1980s [57]. Consistently, a study in Israel found that teenage boys who attended Orthodox schools had much higher rates of myopia than students from other schools who spent less time reading books in the 1990s [53]. Therefore, education and intensive schooling may have a large contribution to the increase in myopia prevalence [58]. Recent studies have also extensively described the relationship between education and visual impairment [59]. Furthermore, a high prevalence of myopia from Taiwan was found in cohorts with low exposure to digital devices [56]. Thus, it is still debatable whether smartphone overuse would lead to a higher risk of myopia or other visual problems.

Second, most of the studies included in this analysis divided smartphone overuse as use time over 2 or 3 hours per day. However, there is some evidence that the time people actually spend engaged with a digital screen is far longer [60-62], suggesting that people may use other electronic devices. Overuse of other digital devices might also play an important role in visual impairment. Some studies have explored the relationships between digital screen time (eg, computer, tablet, smartphone, or other handheld electronic screens) and visual impairment [58,61,63-66]. For instance, a birth cohort study (N=5074 participants) showed that increased computer use was associated with myopia development in children [65]. Yang et al [63] found that screen exposure was significantly and positively associated with preschool myopia, which is consistent with the results of another cohort study [66]. However, the results of studies assessing the impacts of screen time on visual impairment have been mixed. A recent systematic review showed that screen time was not significantly associated with the prevalence and incidence of myopia [58], which may largely support our pooled result of cross-sectional studies. Thus, the relationship needs to be further validated. Moreover, given differences in the use of various digital devices, some studies have compared the impacts of smartphone use with other digital devices on visual impairment [20,24,25,27]. These results are also inconsistent. For instance, Guan et al [47] (N=19,934 participants) found that prolonged (>60 minutes/day) computer usage and smartphone usage were both significantly associated with greater refractive error. Nevertheless, Liu et al [24] and Huang et al [25] found that myopia in children was not associated with time spent using various electronic devices, including smartphones, tablets, and computers. By contrast, a study with a representative sample of 1884 adolescents showed that smartphone use time was associated with an increased risk of visual symptoms, but no significant association was found for tablet use [27]. A controlled trial (N=50 participants) indicated that the smartphone use group had higher fatigue, burning, and dryness scores than the computer use group [20]. Although the existing research supports that smartphone use might cause worse vision than other digital devices, further convincing evidence is needed to support this conclusion owing to the low number of studies.

Third, several studies have shown that technology use or screen time alone is of minimal risk to visual impairment, whereas more time spent outdoors is related to a reduced risk of myopia and myopic progression [25,67,68]. However, there is also evidence that the increased use of digital devices is associated with more time at work and less time spent outdoors, resulting in a substitution effect [58,69]. For example, Dirani et al [69] reported that the lack of adequate outdoor activity might be related to the rise in digital screen time. More specifically, recent educational screen time might be a replacement for reading or writing, in addition to recreational screen time (eg, computer or video games) [69]. For instance, smartphones are used by children mainly for playing games (29%) and watching videos (20%) but also for learning (19%) [70]. Thus, digital screen time might not be a causal factor, but may be a substitute for a different types of work [58]. There is also some evidence that children 9-11 years old who spent less than 2 hours playing on a computer were 1.98 times more likely to spend more than 1 hour outside than those reporting 2 or more hours of computer use [71]. Although these results might reflect a tradeoff between outdoor time and digital screen time, with screen time being a proxy for indoor time, there is no evidence to confirm this substitution effect [58]. Thus, further studies in this field are warranted.

Besides the findings in the cross-sectional studies, we also found that the smartphone overuse group presented worse visual function scores than the reduced-use group in each of the included controlled trials and in the pooled result. Biologically, the effects of smartphones on ocular symptoms can be explained by two types of electromagnetic fields (EMFs): extremely low-frequency EMFs and radiofrequency (RF) electromagnetic radiation (EMR) [72,73]. The intensity of radiation from mobile phones is relatively low with a specific absorption rate <4 W/kg [72,74]. However, it has been reported that adverse effects such as DNA damage and thickening of the cornea occur even at a specific absorption rate lower than 4 W/kg [72,75,76]. The local specific absorption rate has been shown to be higher in tissues at a younger age, suggesting higher susceptibility of adolescents to smartphones [77]. The EMFs generated by smartphones may interact with the tissues of the eyes [73,78], which may cause apoptosis, cataract formation, edema, endothelial cell loss, inflammatory responses, and neurological effects [72,74,79,80]. The RF EMR may affect the body thermally and nonthermally [81], which may result in oxidative stress in the cornea and the lens [74]. These effects by EMFs and RF EMR on the eyes, especially on the cornea and the lens, could suggest why ocular symptoms such as blurring, redness, visual disturbance, inflammation, and lacrimation increase with more exposure to smartphones [23]. Although experimental studies may provide causal inferences, our result needs to be further confirmed due to the limited number of existing studies.

Regarding the association between smartphone overuse and myopia examined in the cross-sectional studies, multiple ocular symptoms found in the experimental studies do not necessarily reflect pathological changes in the eyes, such as myopia. Few longitudinal cohort studies have examined the impacts of screen exposure on myopia, and the results are inconsistent [66,82]. To our knowledge, there have been no experimental or longitudinal studies detecting the impacts of smartphone overuse on myopia specifically. Thus, a longitudinal cohort study design establishing the temporal sequence of prior exposure to environmental factors would be useful to examine whether smartphone overuse may increase the risk of developing myopia.

In addition, the heterogeneity was high in the meta-analysis of included cross-sectional studies. First, a large number of studies have identified potential risk factors that may result in visual impairment, which included both genetic and environmental factors [20,26,56,58] such as age [26], education and occupation [58], outdoor activity [20,58], and parental myopia [20]. However, some studies did not include these variables in the multivariate analysis, which might contribute to the inconsistent findings, and might further affect the individual effect estimates and the pooled OR. Second, some studies only used univariate analysis to investigate the associations between smartphone use time and visual impairment [47,69], which might hinder the exploration of their interrelationships. Third, the assessment of the outcome was inconsistent. For example, some studies used a self-reported questionnaire to identify myopia [26,27], while others used an objective assessment [24,25]. Furthermore, the division of smartphone overuse was inconsistent, which may have precluded us from determining their significant relationships. A guideline advised limiting recreational screen time to no more than 2 hours per day [83]. Therefore, further studies in this field should use a broadly recognized standard to define smartphone overuse.

There are also other limitations of this study that need to be addressed. All of the included studies used a self-reported questionnaire to evaluate smartphone use time. Participants in the included experimental studies also mostly reported their visual function using questionnaires. The questionnaires themselves may be a potential source of error due to inaccurate reporting or recall bias of the participants. Further research should adopt objective instruments to measure smartphone use time and visual acuity screening to examine visual function. Furthermore, generalization of the results should be interpreted with caution owing to the low number of studies included in each meta-analysis. Limiting the review to studies reported in English may have also resulted in nonreporting of studies published in other languages. Nevertheless, our review involved rigorous methodological procedures to obtain and pool data from 27,110 children and young adults. We also adopted a wide range of search terms to retrieve all potential articles published in English, including the grey literature, which might have helped to reduce the publication bias in the final combination.

Conclusions

Overall, current evidence suggests that the results of the association between smartphone overuse and visual impairment in children and young adults are mixed. Although the statistically significantly negative association between smartphone overuse and visual impairment in the meta-analysis was only confirmed in controlled trials and not in cross-sectional studies, the adverse effect of smartphone overuse on visual functions was more apparent in children. However, these relationships need to be further verified. Further research on the patterns of use, with longer follow-up periods to detect longitudinal associations, and the exact mechanisms underlying these associations will help inform detailed guidelines for smartphone use in children and young adults. In addition, understanding the factors of smartphone overuse that account for the risk of ocular symptoms could help the growing population of smartphone users, especially children and young adults, to use smartphones in a healthier manner.

Authors' Contributions

Conceptualization: DZ, ML, and YC; Data curation: JW, ML, and YC; Formal analysis: JW and YC; Investigation: JW, DZ, and ML; Methodology: DZ and YC; Project administration: YC; Software: YC; Supervision: DZ and YC; Validation: JW and YC; Visualization: YC; Writing – original draft: JW and YC; Writing – review & editing: JW, DZ, ML, and YC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategy and results.
[DOCX File , 20 KB - jmir_v22i12e21923_app1.docx]

Multimedia Appendix 2
Tables of study quality assessment.
[DOC File , 69 KB - jmir_v22i12e21923_app2.doc]

References

1. Böhm S. An analysis on country-specific characteristics influencing smartphone usage. 2015 Presented at: First International Conference on Multidisciplinary in Management; 2015; Bangkok p. 335-348.
2. Ertemel A, Ari E. A Marketing Approach to a Psychological Problem: Problematic Smartphone Use on Adolescents. Int J Environ Res Public Health 2020 Apr 04;17(7):2471 [FREE Full text] [doi: 10.3390/ijerph17072471] [Medline: 32260429]
3. Zhai X, Ye M, Wang C, Gu Q, Huang T, Wang K, et al. Associations among physical activity and smartphone use with perceived stress and sleep quality of Chinese college students. Ment Health Phys Act 2020 Mar;18:100323. [doi: 10.1016/j.mhpa.2020.100323]
4. Csibi S, Griffiths MD, Demetrovics Z, Szabo A. Analysis of Problematic Smartphone Use Across Different Age Groups within the ‘Components Model of Addiction’. Int J Ment Health Addiction 2019 May 23:1-16. [doi: 10.1007/s11469-019-00095-0]
5. Li L, Zhong H, Li J, Li C, Pan C. Incidence of myopia and biometric characteristics of premyopic eyes among Chinese children and adolescents. BMC Ophthalmol 2018 Jul 20;18(1):178 [FREE Full text] [doi: 10.1186/s12886-018-0836-9] [Medline: 30029645]
6. Wang SK, Guo Y, Liao C, Chen Y, Su G, Zhang G, et al. Incidence of and Factors Associated With Myopia and High Myopia in Chinese Children, Based on Refraction Without Cycloplegia. JAMA Ophthalmol 2018 Sep 01;136(9):1017-1024 [FREE Full text] [doi: 10.1001/jamaophthalmol.2018.2658] [Medline: 29978185]
7. Grzybowski A, Kanclerz P, Tsubota K, Lanca C, Saw S. A review on the epidemiology of myopia in school children worldwide. BMC Ophthalmol 2020 Jan 14;20(1):27 [FREE Full text] [doi: 10.1186/s12886-019-1220-0] [Medline: 31937276]
8. Li Y, Liu J, Qi P. The increasing prevalence of myopia in junior high school students in the Haidian District of Beijing, China: a 10-year population-based survey. BMC Ophthalmol 2017 Jun 12;17(1):88 [FREE Full text] [doi: 10.1186/s12886-017-0483-6] [Medline: 28606071]
9. Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. Global Prevalence of Myopia and High Myopia and Temporal Trends from 2000 through 2050. Ophthalmology 2016 May;123(5):1036-1042 [FREE Full text] [doi: 10.1016/j.ophtha.2016.01.006] [Medline: 26875007]
10. Dolgin E. The myopia boom. Nature 2015 Mar 19;519(7543):276-278. [doi: 10.1038/519276a] [Medline: 25788077]
11. Choi S, Kim D, Choi J, Ahn H, Choi E, Song W, et al. Comparison of risk and protective factors associated with smartphone addiction and Internet addiction. J Behav Addict 2015 Dec;4(4):308-314 [FREE Full text] [doi: 10.1556/2006.4.2015.043] [Medline: 26690626]
12. Liu C, Lin S, Pan Y, Lin Y. Smartphone gaming and frequent use pattern associated with smartphone addiction. Medicine (Baltimore) 2016 Jul;95(28):e4068 [FREE Full text] [doi: 10.1097/MD.0000000000004068] [Medline: 27428191]
13. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]

14. Bernard JY, Padmapriya N, Chen B, Cai S, Tan KH, Yap F, et al. Predictors of screen viewing time in young Singaporean children: the GUSTO cohort. Int J Behav Nutr Phys Act 2017 Sep 05;14(1):112 [FREE Full text] [doi: 10.1186/s12966-017-0562-3] [Medline: 28870219]
15. LeBlanc AG, Katzmarzyk PT, Barreira TV, Broyles ST, Chaput J, Church TS, ISCOLE Research Group. Correlates of Total Sedentary Time and Screen Time in 9-11 Year-Old Children around the World: The International Study of Childhood Obesity, Lifestyle and the Environment. PLoS One 2015;10(6):e0129622 [FREE Full text] [doi: 10.1371/journal.pone.0129622] [Medline: 26068231]
16. Zou Y, Xia N, Zou Y, Chen Z, Wen Y. Smartphone addiction may be associated with adolescent hypertension: a cross-sectional study among junior school students in China. BMC Pediatr 2019 Sep 04;19(1):310 [FREE Full text] [doi: 10.1186/s12887-019-1699-9] [Medline: 31484568]
17. Lee S, Kang H, Shin G. Head flexion angle while using a smartphone. Ergonomics 2015 Oct 17;58(2):220-226. [doi: 10.1080/00140139.2014.967311] [Medline: 25323467]
18. van Deursen AJ, Bolle CL, Hegner SM, Kommers PA. Modeling habitual and addictive smartphone behavior. Comput Hum Behav 2015 Apr;45:411-420. [doi: 10.1016/j.chb.2014.12.039]
19. Pan C, Qiu Q, Qian D, Hu D, Li J, Saw S, et al. Iris colour in relation to myopia among Chinese school-aged children. Ophthalmic Physiol Opt 2018 Jan 20;38(1):48-55. [doi: 10.1111/opo.12427] [Medline: 29265474]
20. Choi JH, Li Y, Kim SH, Jin R, Kim YH, Choi W, et al. The influences of smartphone use on the status of the tear film and ocular surface. PLoS One 2018 Oct 31;13(10):e0206541 [FREE Full text] [doi: 10.1371/journal.pone.0206541] [Medline: 30379901]
21. Antona B, Barrio AR, Gascó A, Pinar A, González-Pérez M, Puell MC. Symptoms associated with reading from a smartphone in conditions of light and dark. Appl Ergon 2018 Apr;68:12-17. [doi: 10.1016/j.apergo.2017.10.014] [Medline: 29409625]
22. Long J, Cheung R, Duong S, Paynter R, Asper L. Viewing distance and eyestrain symptoms with prolonged viewing of smartphones. Clin Exp Optom 2017 Mar 08;100(2):133-137. [doi: 10.1111/cxo.12453] [Medline: 27716998]
23. Kim J, Hwang Y, Kang S, Kim M, Kim TS, Kim J, et al. Association between Exposure to Smartphones and Ocular Health in Adolescents. Ophthalmic Epidemiol 2016 Aug;23(4):269-276. [doi: 10.3109/09286586.2015.1136652] [Medline: 27254040]
24. Liu S, Ye S, Xi W, Zhang X. Electronic devices and myopic refraction among children aged 6-14 years in urban areas of Tianjin, China. Ophthalmic Physiol Opt 2019 Jul 17;39(4):282-293. [doi: 10.1111/opo.12620] [Medline: 31099434]
25. Huang L, Kawasaki H, Liu Y, Wang Z. The prevalence of myopia and the factors associated with it among university students in Nanjing: A cross-sectional study. Medicine (Baltimore) 2019 Mar;98(10):e14777. [doi: 10.1097/MD.0000000000014777] [Medline: 30855486]
26. McCrann S, Loughman J, Butler JS, Paudel N, Flitcroft DI. Smartphone use as a possible risk factor for myopia. Clin Exp Optom 2020 May 25:online ahead of print. [doi: 10.1111/cxo.13092] [Medline: 32452059]
27. Toh SH, Coenen P, Howie EK, Mukherjee S, Mackey DA, Straker LM. Mobile touch screen device use and associations with musculoskeletal symptoms and visual health in a nationally representative sample of Singaporean adolescents. Ergonomics 2019 Jun 23;62(6):778-793. [doi: 10.1080/00140139.2018.1562107] [Medline: 30575442]
28. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
29. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000 Apr 19;283(15):2008-2012. [doi: 10.1001/jama.283.15.2008] [Medline: 10789670]
30. VHL Regional Portal. URL: https://bvsalud.org/en/ [accessed 2020-11-18]
31. NARCIS. Data Archiving and Networked Services (DANS). URL: https://www.narcis.nl [accessed 2020-11-18]
32. Grey Literature Report. The New York Academy of Medicine. URL: http://greylit.org [accessed 2020-11-18]
33. Open Grey. URL: http://opengrey.eu [accessed 2020-11-18]
34. Erikson E, Erikson J. The life cycle completed (extended version). New York: WW Norton & Company; 1998.
35. Zastrow C, Kirst-Ashman K. Understanding human behavior and the social environment. Hampshire: Cengage Learning; 2006.
36. Critical Appraisal Tools. Joanna Briggs Institute. 2017. URL: https://joannabriggs.org/critical-appraisal-tools [accessed 2020-11-28]
37. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 21;278(12):652-657. [doi: 10.1056/NEJM196803212781204] [Medline: 5637250]
38. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
39. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions Version 5.1. London: The Cochrane Collaboration; 2018.
40. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]

41. Baujat B, Mahé C, Pignon J, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002 Sep 30;21(18):2641-2652. [doi: 10.1002/sim.1221] [Medline: 12228882]
42. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
43. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
44. Tobias A. Assessing the influence of a single study in the meta-analysis estimate. Stata Tech Bull 1999;8(47):15-17.
45. Schwarzer G, Carpenter J, Rücker G. Meta-analysis with R. Berlin: Springer; 2015.
46. Harrer M, Cuijpers P, Furukawa T, Ebert DD. dmetar: Companion R package for the Guide ‘Doing meta-analysis is R’. URL: https://dmetar.protectlab.org/reference/dmetar.html [accessed 2020-11-28]
47. Guan H, Yu NN, Wang H, Boswell M, Shi Y, Rozelle S, et al. Impact of various types of near work and time spent outdoors at different times of day on visual acuity and refractive error among Chinese school-going children. PLoS One 2019 Apr 26;14(4):e0215827 [FREE Full text] [doi: 10.1371/journal.pone.0215827] [Medline: 31026279]
48. Alharbi M, Alem A, Alrizgi H, Abualnassr K, Abualnassr F, Alshencieti A. Impact of Smartphones on Eye Health among Health Sciences Students in Taibah University, Saudi Arabia. Indo Am J Pharm Sci 2019;6(2):3523-3530. [doi: 10.5281/zenodo.2561230]
49. Küçer N. Some ocular symptoms experienced by users of mobile phones. Electromagn Biol Med 2008 Jul 07;27(2):205-209. [doi: 10.1080/15368370802072174] [Medline: 18568938]
50. Meo SA, Al-Drees AM. Mobile phone related-hazards and subjective hearing and vision symptoms in the Saudi population. Int J Occup Med Environ Health 2005;18(1):53-57. [Medline: 16052891]
51. Merrie YA, Tegegne MM, Munaw MB, Alemu HW. Prevalence And Associated Factors Of Visual Impairment Among School-Age Children In Bahir Dar City, Northwest Ethiopia. Clin Optom (Auckl) 2019;11:135-143. [doi: 10.2147/OPTO.S213047] [Medline: 31807106]
52. Lee D, Hong S, Jung S, Lee K, Lee G. The Effects of Viewing Smart Devices on Static Balance, Oculomotor Function, and Dizziness in Healthy Adults. Med Sci Monit 2019 Oct 27;25:8055-8060 [FREE Full text] [doi: 10.12659/MSM.915284] [Medline: 31655845]
53. Zylbermann R, Landau D, Berson D. The influence of study habits on myopia in Jewish teenagers. J Pediatr Ophthalmol Strabismus 1993;30(5):319-322. [Medline: 8254449]
54. Lin LLK, Shih YF, Hsiao CK, Chen CJ. Prevalence of myopia in Taiwanese schoolchildren: 1983 to 2000. Ann Acad Med Singap 2004 Jan;33(1):27-33 [FREE Full text] [Medline: 15008558]
55. Morgan I, Rose K. How genetic is school myopia? Prog Retin Eye Res 2005 Jan;24(1):1-38. [doi: 10.1016/j.preteyeres.2004.06.004] [Medline: 15555525]
56. Morgan IG, French AN, Ashby RS, Guo X, Ding X, He M, et al. The epidemics of myopia: Aetiology and prevention. Prog Retin Eye Res 2018 Jan;62:134-149. [doi: 10.1016/j.preteyeres.2017.09.004] [Medline: 28951126]
57. Sensaki S, Sabanayagam C, Verkicharla PK, Awodele A, Tan KH, Chia A, et al. An Ecologic Study of Trends in the Prevalence of Myopia in Chinese Adults in Singapore Born from the 1920s to 1980s. Ann Acad Med Singap 2017 Jun;46(6):229-236 [FREE Full text] [Medline: 28733687]
58. Lanca C, Saw S. The association between digital screen time and myopia: A systematic review. Ophthalmic Physiol Opt 2020 Mar 13;40(2):216-229. [doi: 10.1111/opo.12657] [Medline: 31943280]
59. Morgan IG, Rose KA. Myopia: is the nature-nurture debate finally over? Clin Exp Optom 2019 Jan;102(1):3-17. [doi: 10.1111/cxo.12845] [Medline: 30380590]
60. Hansen MH, Laigaard PP, Olsen EM, Skovgaard AM, Larsen M, Kessel L, et al. Low physical activity and higher use of screen devices are associated with myopia at the age of 16-17 years in the CCC2000 Eye Study. Acta Ophthalmol 2020 May;98(3):315-321. [doi: 10.1111/aos.14242] [Medline: 31502414]
61. Gopinath B, Hardy LL, Baur LA, Teber E, Mitchell P. Influence of parental history of hypertension on screen time and physical activity in young offspring. J Hypertens 2012 Feb;30(2):336-341. [doi: 10.1097/HJH.0b013e32834ea436] [Medline: 22179085]
62. Harrington SC, Stack J, O'Dwyer V. Risk factors associated with myopia in schoolchildren in Ireland. Br J Ophthalmol 2019 Dec;103(12):1803-1809. [doi: 10.1136/bjophthalmol-2018-313325] [Medline: 30745305]
63. Yang G, Huang L, Schmid KL, Li C, Chen J, He G, et al. Associations Between Screen Exposure in Early Life and Myopia amongst Chinese Preschoolers. Int J Environ Res Public Health 2020 Feb 07;17(3):1056 [FREE Full text] [doi: 10.3390/ijerph17031056] [Medline: 32046062]
64. Toh SH, Coenen P, Howie EK, Smith AJ, Mukherjee S, Mackey DA, et al. A prospective longitudinal study of mobile touch screen device use and musculoskeletal symptoms and visual health in adolescents. Appl Ergon 2020 May;85:103028. [doi: 10.1016/j.apergo.2019.103028] [Medline: 32174368]
65. Enthoven CA, Tideman JWL, Polling JR, Yang-Huang J, Raat H, Klaver CCW. The impact of computer use on myopia development in childhood: The Generation R study. Prev Med 2020 Mar;132:105988 [FREE Full text] [doi: 10.1016/j.ypmed.2020.105988] [Medline: 31954142]

66. Jones-Jordan LA, Mitchell GL, Cotter SA, Kleinstein RN, Manny RE, Mutti DO, CLEERE Study Group. Visual activity before and after the onset of juvenile myopia. Invest Ophthalmol Vis Sci 2011 Mar;52(3):1841-1850 [FREE Full text] [doi: 10.1167/iovs.09-4997] [Medline: 20926821]
67. Wu P, Tsai C, Wu H, Yang Y, Kuo H. Outdoor activity during class recess reduces myopia onset and progression in school children. Ophthalmology 2013 May;120(5):1080-1085. [doi: 10.1016/j.ophtha.2012.11.009] [Medline: 23462271]
68. Ramamurthy D, Lin Chua SY, Saw S. A review of environmental risk factors for myopia during early life, childhood and adolescence. Clin Exp Optom 2015 Nov;98(6):497-506. [doi: 10.1111/cxo.12346] [Medline: 26497977]
69. Dirani M, Crowston JG, Wong TY. From reading books to increased smart device screen time. Br J Ophthalmol 2019 Jan;103(1):1-2. [doi: 10.1136/bjophthalmol-2018-313295] [Medline: 30401675]
70. McCloskey M, Johnson SL, Benz C, Thompson DA, Chamberlin B, Clark L, et al. Parent Perceptions of Mobile Device Use Among Preschool-Aged Children in Rural Head Start Centers. J Nutr Educ Behav 2018 Jan;50(1):83-89. [doi: 10.1016/j.jneb.2017.03.006] [Medline: 29031581]
71. Wilkie HJ, Standage M, Gillison FB, Cumming SP, Katzmarzyk PT. The home electronic media environment and parental safety concerns: relationships with outdoor time after school and over the weekend among 9-11 year old children. BMC Public Health 2018 Apr 05;18(1):456 [FREE Full text] [doi: 10.1186/s12889-018-5382-0] [Medline: 29621981]
72. Hardell L, Sage C. Biological effects from electromagnetic field exposure and public exposure standards. Biomed Pharmacother 2008 Feb;62(2):104-109. [doi: 10.1016/j.biopha.2007.12.004] [Medline: 18242044]
73. Kheifets L, Repacholi M, Saunders R, van Deventer E. The sensitivity of children to electromagnetic fields. Pediatrics 2005 Aug;116(2):e303-e313. [doi: 10.1542/peds.2004-2541] [Medline: 16061584]
74. Balci M, Devrim E, Durak I. Effects of mobile phones on oxidant/antioxidant balance in cornea and lens of rats. Curr Eye Res 2007 Jan;32(1):21-25. [doi: 10.1080/02713680601114948] [Medline: 17364731]
75. Akar A, Karayiğit MO, Bolat D, Gültiken ME, Yarim M, Castellani G. Effects of low level electromagnetic field exposure at 2.45 GHz on rat cornea. Int J Radiat Biol 2013 Apr;89(4):243-249. [doi: 10.3109/09553002.2013.754557] [Medline: 23206266]
76. Lixia S, Yao K, Kaijun W, Deqiang L, Huajun H, Xiangwei G, et al. Effects of 1.8 GHz radiofrequency field on DNA damage and expression of heat shock protein 70 in human lens epithelial cells. Mutat Res 2006 Dec 01;602(1-2):135-142. [doi: 10.1016/j.mrfmmm.2006.08.010] [Medline: 17011595]
77. Leitgeb N. Mobile phones: are children at higher risk? Wien Med Wochenschr 2008;158(1-2):36-41. [doi: 10.1007/s10354-007-0447-1] [Medline: 18286248]
78. Khurana VG, Teo C, Kundi M, Hardell L, Carlberg M. Cell phones and brain tumors: a review including the long-term epidemiologic data. Surg Neurol 2009 Sep;72(3):205-214; discussion 214. [doi: 10.1016/j.surneu.2009.01.019] [Medline: 19328536]
79. Elder JA. Ocular effects of radiofrequency energy. Bioelectromagnetics 2003;Suppl 6:S148-S161. [doi: 10.1002/bem.10117] [Medline: 14628311]
80. Yao K, Wang KJ, Sun ZH, Tan J, Xu W, Zhu LJ, et al. Low power microwave radiation inhibits the proliferation of rabbit lens epithelial cells by upregulating P27Kip1 expression. Mol Vis 2004 Feb 25;10:138-143 [FREE Full text] [Medline: 14990889]
81. Challis LJ. Mechanisms for interaction between RF fields and biological tissue. Bioelectromagnetics 2005;Suppl 7:S98-S106. [doi: 10.1002/bem.20119] [Medline: 15931683]
82. Jones LA, Sinnott LT, Mutti DO, Mitchell GL, Moeschberger ML, Zadnik K. Parental history of myopia, sports and outdoor activities, and future myopia. Invest Ophthalmol Vis Sci 2007 Aug;48(8):3524-3532 [FREE Full text] [doi: 10.1167/iovs.06-1118] [Medline: 17652719]
83. Rose KA, Morgan IG, Ip J, Kifley A, Huynh S, Smith W, et al. Outdoor activity reduces the prevalence of myopia in children. Ophthalmology 2008 Aug;115(8):1279-1285. [doi: 10.1016/j.ophtha.2007.12.019] [Medline: 18294691]

Abbreviations

EMF: electromagnetic field
EMR: electromagnetic radiation
ES: effect size
JBI: Joanna Briggs Institute
LOO: leave-one-out
OR: odds ratio
RF: radiofrequency

Edited by G Eysenbach; submitted 29.06.20; peer-reviewed by K Turner, T Igbe; comments to author 26.07.20; accepted 29.10.20; published 08.12.20.

Please cite as:
Wang J, Li M, Zhu D, Cao Y
Smartphone Overuse and Visual Impairment in Children and Young Adults: Systematic Review and Meta-Analysis
J Med Internet Res 2020;22(12):e21923
URL: https://www.jmir.org/2020/12/e21923
doi:10.2196/21923
PMID:33289673

Review

Design Features for Improving Mobile Health Intervention User Engagement: Systematic Review and Thematic Analysis

Yanxia Wei[1], PhD; Pinpin Zheng[1], PhD; Hui Deng[1], BA; Xihui Wang[1], MA; Xiaomei Li[1], BA; Hua Fu[1], PhD

[1] School of Public Health, Fudan University, Shanghai, China

Corresponding Author:
Pinpin Zheng, PhD
School of Public Health
Fudan University
138 Yixueyuan Road
Shanghai, 200030
China
Phone: 86 021 54237202
Email: zpinpin@shmu.edu.cn

Abstract

Background: Well-designed mobile health (mHealth) interventions support a positive user experience; however, a high rate of disengagement has been reported as a common concern regarding mHealth interventions. To address this issue, it is necessary to summarize the design features that improve user engagement based on research over the past 10 years, during which time the popularity of mHealth interventions has rapidly increased due to the use of smartphones.

Objective: The aim of this review was to answer the question “Which design features improve user engagement with mHealth interventions?” by summarizing published literature with the purpose of guiding the design of future mHealth interventions.

Methods: This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist. Databases, namely, PubMed, Web of Science, Cochrane Library, Ovid EMBASE, and Ovid PsycINFO, were searched for English and Chinese language papers published from January 2009 to June 2019. Thematic analysis was undertaken to assess the design features in eligible studies. The Mixed Methods Appraisal Tool was used to assess study quality.

Results: A total of 35 articles were included. The investigated mHealth interventions were mainly used in unhealthy lifestyle (n=17) and chronic disease (n=10) prevention programs. Mobile phone apps (n=24) were the most common delivery method. Qualitative (n=22) and mixed methods (n=9) designs were widely represented. We identified the following 7 themes that influenced user engagement: personalization (n=29), reinforcement (n=23), communication (n=20), navigation (n=17), credibility (n=16), message presentation (n=16), and interface aesthetics (n=7). A checklist was developed that contained these 7 design features and 29 corresponding specific implementations derived from the studies.

Conclusions: This systematic review and thematic synthesis identified useful design features that make an mHealth intervention more user friendly. We generated a checklist with evidence-based items to enable developers to use our findings easily. Future evaluations should use more robust quantitative approaches to elucidate the relationships between design features and user engagement.

(J Med Internet Res 2020;22(12):e21687) doi:10.2196/21687

KEYWORDS

mHealth; design feature; user engagement; thematic synthesis analysis

Introduction

Mobile health (mHealth) is a means of providing medical and public health support to health care consumers via mobile devices, such as mobile phones, portable computers, and personal digital assistants [1]. mHealth interventions involve the adoption of mobile technologies to provide educational information, help users manage their own conditions and behaviors, and deliver health care to improve the health of users. Compared to traditional delivery models, mobile interventions can be more cost-effective [2,3]; help users overcome demographic, socioeconomic, and geographic barriers to access [4-6]; allow the privacy of users to be protected [7]; and allow a high level of customization, self-management, and communication [8-11]. With the popularity of smartphones, mHealth technology has grown rapidly in the past 10 years and has been used in many health fields. Mobile interventions have been shown to improve healthy behaviors (eg, weight loss [12], balanced diet consumption [13], and smoking cessation [14]) and disease management [15] (eg, tuberculosis [16] and AIDS [17]).

mHealth programs require autonomous use [5], which depends on user involvement and self-management, and engagement is related to behavior changes and health improvements [18]. User engagement refers to high uptake, high-quality user experience, and good adherence over long periods of time [19]. However, a pressing concern regarding mHealth interventions is the high reported rate of disengagement [4]. For example, in a web-based weight loss study, respondents did not use the app as intended, and only 64% of the intervention group actually used the intervention at least once [20]. A mobile phone text message smoking cessation program also suffered from a high attrition rate: nearly half of the subscribers did not complete the entire program in the real-world implementation phase, and the majority of opt-outs occurred in the first 2 weeks [21]. With regard to web-based physical activity interventions, reported dropout attrition rates vary between 0% and 62% [22,23]. This is a common issue related to mHealth interventions; poor user engagement makes intervention effectiveness difficult because the users are not exposed to enough of the intervention content [24].

The development steps of an mHealth intervention tool can influence user engagement. The first step is intervention content development. User engagement will be low if the content does not adhere to what has been shown to be effective [25,26], and there are already well-established guidelines for health intervention content development [25,27,28]. The second step is design of the mode by which the intervention content is delivered (eg, information architecture, screen appearance, and interactive features). Poor design features, such as complicated navigation and difficult-to-read screen presentations, are poorly tolerated by users in real-world settings [29]. Good mHealth design is readily distinguishable from its competitors, leading users to feel more favorably disposed toward the product and have a positive user experience [4,19,30,31]. Several studies analyzed which design features should be included in mHealth interventions [32-35]; however, they did not clearly describe how to specifically deliver the interventions, and the contributions of most of the design features mentioned have not been tested in empirical studies. In addition, Morrison et al [28] developed a hypothetical framework to define the design features through a review, but the framework only focused on 4 interactive design features (social context and support, contact with the intervention, tailoring, and self-management) and provided simple definitions of the other 8 features that were difficult to implement in subsequent studies. Webb et al [36] developed a coding scheme for design features in a meta-analysis; however, this scheme was proposed before a review of the literature was performed rather than derived from the literature reviewed, so it inevitably missed features that are important but not reported.

While these studies have provided some important guidance for the design features of mHealth interventions, none has included comprehensive design features based on the literature or experiences. Additionally, except for Crutzen et al [34], other researchers did not focus on the relationships between the design features mentioned and user engagement. Furthermore, the data on which these studies were based were obtained more than 10 years ago, and in the past 10 years, the popularity of mHealth interventions has rapidly expanded due to the use of smartphones; data from more recent studies need to be considered. The objectives of our study were to systematically review studies published in the past 10 years regarding design features that improve user engagement with mHealth interventions and generate a checklist that can easily be used during the design of future mHealth interventions.

Methods

Protocol

This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [37]. The protocol was registered in the International Prospective Register of Systematic Reviews (CRD42020140282).

Inclusion and Exclusion Criteria

Articles were selected if they met the following criteria: (1) The study was empirical. The study population was composed of users or potential users of mHealth. If a study focused on special populations, such as children and older adults, it was excluded. (2) The study focused on the mode of delivery of health intervention content via mobile devices, for example, aesthetics, message phrasing, and interactive features. Articles that mentioned design features but did not explain them, making it unclear which features were being referenced, were excluded. Studies on the selection of intervention content or theory, and those providing a general description of the process of designing an mHealth intervention were excluded. (3) The study reported quantitative and qualitative analyses of the effectiveness of the design features with regard to increasing user engagement or user acceptance of the design features. (4) The article was published in a peer-reviewed journal from January 1, 2009 to June 13, 2019. (5) The article was published in English or Chinese.

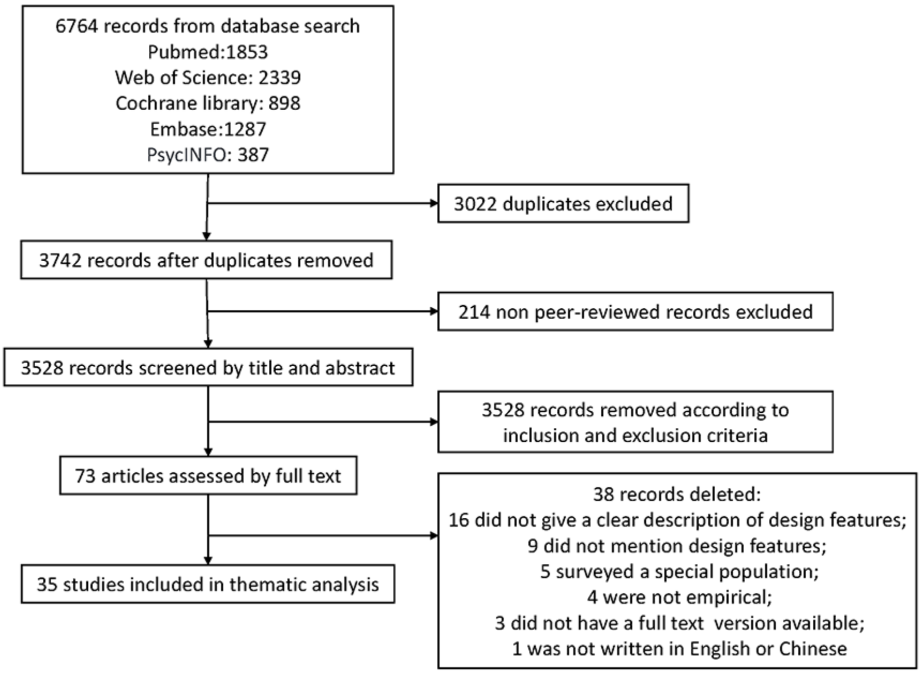
Search and Screening Strategy

Five databases, namely, PubMed, Web of Science, Cochrane Library, Ovid EMBASE, and Ovid PsycINFO, were searched in June 2019. The search strategy was *"mobile health"* or *mHealth* or *m-health* or *eHealth* or *e-health* combined with *design or feature* or principle** or *"mode* of delivery"* or *model* and combined with *engage** or *adhere** or *maintain** or *retention or sustain** or *usage** or *satisf** or *prefer* or *preference** or *accept** or *reliable*.

The search results were uploaded to EndNote (Version X9; Clarivate Analytics) for screening. Figure 1 shows the process of identifying the eligible articles. Duplicates were identified with the sorting function in EndNote. Peer-reviewed journal articles were checked by searching Ulrich's Periodicals Directory; some journals could not be found there, so we visited the journal website to find evidence of the peer-review process.

All remaining articles were assigned to 1 of 3 reviewers and were screened by titles, abstracts, and then full texts according to inclusion and exclusion criteria. YW screened all journal articles independently again. Any disagreement was discussed among the reviewers.

Figure 1. Screening flowchart.



Data Extraction and Synthesis

All selected articles were imported into NVivo (version 11; QSR International). The following study characteristics were extracted: (1) article characteristics, including country and year of publication; (2) health topic; (3) participant characteristics, including sampling methods, sample size, sex, and age; (4) data collection method; and (5) mobile technology. Study characteristics were analyzed using descriptive statistics.

The thematic synthesis analysis method developed by Thomas and Harden [38] has 3 steps: (1) line-by-line coding of the articles to record the components, (2) the development of descriptive themes, and (3) the creation of analytical themes. We performed the first 2 steps together. YW coded each line of the records according to its meaning, translated the concepts among the records, and developed the descriptive themes. HD validated the results by comparing each assigned code to the full texts of the articles. Then, analytical themes were developed by answering the review question using the existing descriptive themes. Each reviewer did this independently, and the results were discussed among all authors. The coding process was iterative.

Assessing Study Quality

The Mixed Methods Appraisal Tool (MMAT) [39] was used to assess the methodological quality of the included studies, including the data collection methods, participant sampling, interpretation of results, consideration of confounders, and risk of bias. In the MMAT, there are 5 criteria with responses “yes,” “no,” and “cannot tell” for each research design. The retained studies were assessed, and the results recorded by 2 authors (YW and HD), independently.

Results

Study Characteristics

All 35 articles included in the analysis were published between 2011 and 2019. They were primarily from the United States (n=14), the United Kingdom (n=9), and Australia (n=5). In terms of the health topics discussed in these studies, 17 articles focused on unhealthy lifestyles (eg, smoking, excessive alcohol consumption, sleep disturbances, and poor sexual health), 10 studies focused on chronic diseases (eg, diabetes, breast cancer, chronic arthritis, and asthma), 4 studies focused on mental health problems, and 4 studies focused on other health issues. Qualitative studies (n=22) and mixed methods (n=9) were the most common data collection methods used, accounting for 89% (31/35) of all studies. The sample sizes ranged from 8 to 1865 in the 35 studies, and the age of the participants ranged from 14 to 74 years old. The mobile technology used in the majority of studies was a mobile phone app (n=24), followed by a website platform (n=6), and text messages (n=5). More detailed information is presented in Multimedia Appendix 1.

Study Quality

A total of 3 articles meet the criteria of all 5 items, with the remaining articles meeting 4 criteria (n=18), 3 criteria (n=13), and 2 criteria (n=1). The most common reasons low scores in each research design were a lack of coherence between the data collection and analysis and the explanations in qualitative studies; the quality of different components was low in the mixed methods studies; and there were poor sampling strategies and a high risk of bias in the quantitative studies (Multimedia Appendix 2).

Themes of Design Features

In total, 7 analytical themes were generated to describe the design features that can improve user engagement with mHealth interventions, and each of these can be explained by several descriptive themes. With regard to the 3 types of mobile technology used in the studies, *mobile phone apps* and *website platforms* share common design features, while *text messages* lack 2 analytical themes: interface aesthetic and reinforcement (Figure 2). For health topics, *unhealthy lifestyles*, *chronic diseases*, and *mental health problems* share common design features, and the topic of *other health issues* lacks interface aesthetic and message presentation (Figure 3).

Figure 2. Comparison of analytical themes in different mobile technologies. TM: text message; WP: website platform; MPA: mobile phone app.

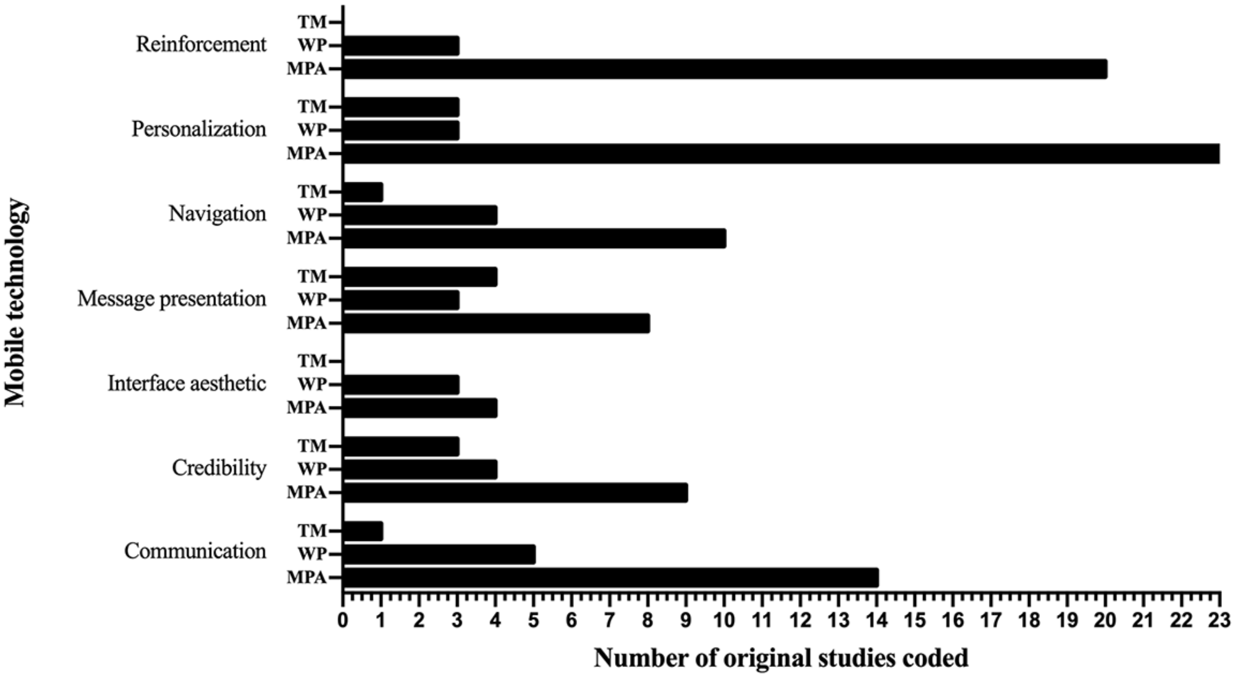
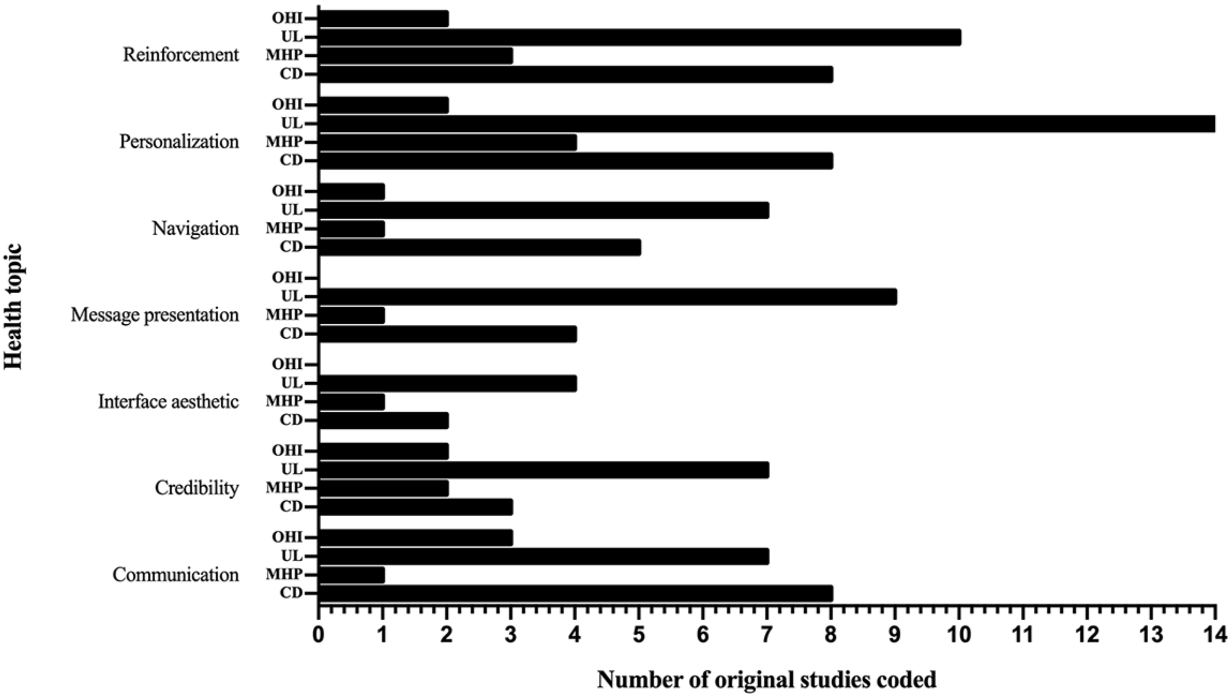


Figure 3. Comparison of analytical themes in different health topics. OHI: other health issues; UL: unhealthy lifestyle; MHP: mental health problem; CD: chronic disease.



Interface Aesthetic

Overview

The interface refers to the appearance of the screen, which was reported to directly impact the user's impression of the intervention and impact user engagement in 7 of the 35 studies (20%). Two descriptive themes that could improve interface aesthetics were identified: (1) attention-grabbing and (2) simple and consistent style.

Attention Grabbing

An aesthetically appealing screen easily attracts user attention [40]. Users preferred that the screen show graphics rather than too much text because the latter can be overwhelming [41,42]. They also preferred a pleasing color scheme. Bright colors (eg, light green, white) were considered attention grabbing, while dark and neon colors discouraged further use [43,44].

Simple and Consistent Style

A simple and clean screen was praised most frequently in the studies, while users disliked complex and overcrowded pages, which rapidly made them lose interest [40,41,44,45]. Many users appreciated the use of a consistent style, with a coherent presentation in terms of colors, pictures, and themes throughout an mHealth intervention [41,46].

Navigation

Overview

Navigation describes how users move to different areas of content within mHealth intervention apps. There were 2 descriptive themes pertaining to navigation: (1) ease of use and (2) automation; these themes were proposed by users in 17 of the 35 studies (49%).

Easy to Use

The users highlighted the importance of minimum input and efficient access to information, such as a simple log-in process, fewer required tasks, or fewer buttons on the screen [43-53]. An interactive process that confused the users or that took them many iterations to understand prevented them from continuing to engage with the app. Explanation of how the mHealth intervention worked (ie, clarifying what to do next) promoted continued use [41,50,53,54].

Automation

Users did not want to spend much time scrolling to find the information they wanted. The search bar and menu bar, which provide options to the user, were thought to facilitate usage [44,46,55-58].

Personalization

Overview

Personalization is a design feature that makes mobile technology act in a particular way depending on user preferences. Personalization was achieved by the following 3 elements according to 29 of the 35 studies (83%): (1) assessment, (2) feedback, and (3) manipulation.

Assessment

Users expected to be assessed with metrics pertaining to the health problem that was the focus of the intervention in as much detail as possible to create an accurate profile [49,59-61]. Some parameters, such as sociodemographic characteristics, basic health status, and individual preferences and habits, could be measured with a quick survey within an app [54,55,57,62,63]. Users also liked continuous monitoring features that allowed them to record their progress toward their goal on a daily basis (eg, health and behavior changes and adherence to an intervention) [40,42,43,45,58,64-67] or diary entries or notes that helped them track their progress [40,45,60]. Moreover, users also indicated a preference for sensor-based automated tracking as opposed to self-reported data, which they often forgot to input and found were not as convenient [47,52,59,63,68].

Feedback

Studies indicated the importance of building an assessment on the basis of feedback on the acquired data; users quickly lost interest when they did not receive feedback that was customized [41,69]. The preference was for the mobile device to provide personalized information, including tailored intervention content matched to their basic characteristics [40,43,45,46,51,54,55,57-65,67,68,70,71] and feedback on continuous monitoring data, for example, their health and behavior progress over time, predicted possible causes and consequences of a health problem and advice on the behavior under investigation [41-43,47,49,52,53,58,59,63,68]. There was a strong interest among users for visualization of continuous monitoring data, for example, presenting data as graphs and tables [47,49,53,58,63,67,69].

Manipulation

Users not only wanted to obtain automatically tailored information but also wanted to be able to customize the mHealth intervention themselves. Users highlighted the importance of being able to choose when and how they receive reminders [43,45,53,57,63-65,72], set goals for the future use of the tool [43,47,49,53,54,58,63,65], and select preferred styles, such as the color and font [46,54,59].

Reinforcement

Overview

Reinforcement is the provision of a stimulus to strengthen the likelihood of a user continuing to exhibit a certain behavior in the future. There were 2 descriptive themes extracted from 23 of the 35 studies (66%), namely, (1) rewards and (2) reminders, that helped enhance reinforcement.

Rewards

The reward feature could increase user motivation to engage with the intervention, and users expressed desire for confirmation when they completed a task. The reward features extracted from the studies included material incentives (eg, cash or gifts), intangible rewards (eg, virtual badges, rankings, certificates, and points), and messages of congratulations when a task was completed [40,41,43,49,51-54,59-61,63].

Reminders

Users wanted reminders to schedule a task, such as taking medicine, making and keeping clinic appointments, and continuing a health plan [49,55,58,60,63,66,68]; to return to their mobile device [42,46,64,70]; and to motivate them with information and advice [40,45,52,61,67,72]. The preferred forms of reminders included email messages, text messages, words of the day, and pop-ups [45,49,52,61,70].

Communication

Overview

Communication is a function that prompts users to consult and communicate with other people regarding their health problems via a mobile device. (1) Peer-to-peer communication and (2) access to professionals were 2 aspects of communication reported in 20 of the 35 studies (57%).

Peer-to-Peer Communication

The users expressed interest in communicating with other people with similar experiences through online forums, communities, by instant messages within an mHealth tool, or by connecting via other types of social media; they wanted to be able to post information, share their stories, ask and answer questions, and find mutual accountability partners [40,45,46,54-56,58-60,63,66,67,69,72].

Access to Professionals

Users wanted to be able to directly contact a health care provider via email, text message, or live chat to ask them questions or obtain advice based on their health data [40,45,48,53-56,58-61,64,66,68,72].

Message Presentation

Overview

The presentation of information is an important factor that impacts user engagement, and information that is presented well is readily accepted by users. (1) Language, (2) tone of voice, and (3) presentation design are points of consideration when seeking to improve message presentation, according to the results of 16 of the 35 studies (46%).

Language

The message needs to be clearly presented in the language used by the audience. The users recommended using simple nontechnical language that is straightforward and concise; they were tired of patronizing and technical language [48,55,62,69,73,74]. When providing an actionable message, users preferred a specific description that made it clear what they were supposed to do [70,71,73].

Tone of Voice

Users recommended using language that framed the information positively rather than negatively, as the latter made them feel discouraged and made them want to turn off the device [43,51,54,62,71,73,74]. To increase user acceptance, it was essential to make the user feel supported and relaxed by using a nonauthoritarian, friendly, and nonjudgmental tone of voice [41,45,69,71].

Presentation Design

The users were quickly bored with text-heavy presentations of information; they wanted multimedia messages, for example, text combined with relevant pictures or video [44,46,53,74]. Knowledge quizzes and games were also recommended as ways to deliver information that prompted the user to engage and learn the information [44,54]. The use of various font styles, sizes, and colors to highlight key information was suggested, as it allowed the user to skim quickly when they lacked the patience to read the entire message [44,46]. Editing the text to make it as concise as possible was also suggested by users, who were not inclined to read lengthy messages [53,62,71].

Credibility

Overview

Credibility is an important feature that guarantees the level of user comfort, enabling them to engage with the mobile technology without experiencing concerns. (1) Trustworthiness and (2) confidentiality were 2 descriptive themes related to credibility derived from 16 of the 35 studies (46%).

Trustworthiness

Users trusted mHealth interventions from authoritative and familiar organizations or developers that were free from advertisements [41,43,48,54,61,63,68,74]. Users emphasized the fact that the information provided needed to be evidence-based and from credible sources to gain their trust [50,57,61,62,65,69].

Confidentiality

Users highlighted the importance of having a privacy policy, for example, a policy that allowed the users to decide whether others could access their data [50,56], ensured that the users remained anonymous when sharing their data with the health care providers or for research [43,64], and allowed users to set passwords for protection [45,61].

Checklist of Design Features to Enhance User Engagement

Based on these themes, we produced a checklist that considers 7 aspects of design and the corresponding implementations based on an exhaustive analysis of the 35 studies. In total, there were 29 items reported that enhance user engagement, and we provide here the descriptions and examples as a reference for future studies (Table 1).

Table 1. Checklist of design features that enhance user engagement.

Items	Criteria
Interface aesthetic	
1	The screen shows a graphic presentation rather than too much information
2	Pleasing color scheme with bright colors (eg, light green, white)
3	Simple screen presentation that is not overcrowded
4	Coherent scheme of colors, pictures, and themes throughout the intervention
Navigation	
5	Minimum user input needed; efficient access to the information provided, such as in a simple menu; and few buttons on the screen
6	Guidance provided that explains how the mHealth intervention works
7	Search bar or menu bar provided to accelerate the process of finding certain information
Personalization	
8	Assessment of the preferences, sociodemographic characteristics and health status of the user
9	Continuous monitoring of health and behavior changes or adherence to an intervention
10	Provision of a diary or note-taking function
11	Provision of personalized information matched to the user's characteristics
12	Provision of feedback on the continuously monitored data
13	Visual presentation of feedback, such as in graphs and tables
14	Provision of autonomy to customize the intervention, for example, allowing the users to choose when and how they receive reminders, to set a goal about their future use of the intervention tool, and to select their preferred styles, such as their preferred colors and fonts
Reinforcement	
15	Provision of material incentives (eg, cash or gifts), intangible rewards (eg, virtual badges, rankings, certificates, and points), or messages of congratulations when a task is completed
16	Sending of reminders to facilitate the scheduling of tasks and to ensure continuous use
Communication	
17	Provision of access to other people with similar experiences through an online forum, community, or instant messages within the mobile tool or by connection with other forms of social media
18	Provision of access to a health care provider through email, text message, or live chat
Message presentation	
19	Use of simple nontechnical language that can be readily understood
20	Use of specific descriptions when providing actionable message
21	Use of a positive, nonauthoritarian, and nonjudgmental tone of voice
22	Provision of multimedia messages, for example, text combined with relevant pictures or videos
23	Presentation of information in the form of knowledge quizzes and games, if possible
24	Use of various font styles, sizes, and colors to highlight information
25	Editing of the text to make it as concise as possible
Credibility	
26	Absence of advertisements
27	Provision of evidence-based information from credible sources
28	Provision of a privacy policy that gives users the right to decide whether others can access their data and ensures the users remain anonymous when sharing their data with the health care providers or for research
29	Enabling users to set a password or code to protection their data

Discussion

Overall Findings

We used thematic synthesis to identify design features that increased user engagement with mHealth interventions based on user feedback. For each design feature, specific implementations in mobile tools were also analyzed. This study presented 7 analytical themes and 16 descriptive themes pertaining to design features that can improve user engagement with mHealth interventions. From most to least commonly mentioned in the studies, the analytical themes were personalization (29/35, 83%), reinforcement (23/35, 66%), communication (20/35, 57%), navigation (17/35, 49%), credibility (16/35, 46%), message presentation (16/35, 46%), and interface aesthetic (7/35, 20%); each analytical theme involves several descriptive subthemes that explain how to implement them when designing mHealth interventions. Overall, the 7 analytical themes were applicable to different mobile technologies and health topics, indicating that the design features identified by this study are universal across mobile apps, website platforms, text messages, and different health themes.

To promote better application of the results of this study to future mHealth intervention development, we developed a checklist of the design features that enhance user engagement; this tool has 29 evidence-based items that are clearly described to make them easy to use by developers of mHealth interventions.

Personalization, reinforcement, and communication were the design features that were mentioned the most often. Compared to other analytical themes, these 3 design features focused on the interactivity of mHealth interventions, including user-to-technology interactions and user-to-user interactions. User-to-technology communication refers to having the user input information about themselves to which the tool provides a tailored response [75]. Two analytical themes, personalization and reinforcement, pertain to the interaction between users and technology. User-to-user interactions are represented by the theme of communication in this article, including peer-to-peer contact and consultation with professionals. Interactive features give users a sense of ownership [76] and promote their participation in the mHealth intervention. Compared with traditional smoking cessation methods, a major value of mobile health interventions is that they can provide better and faster interactions [77] to meet the needs of users.

Personalization, reinforcement, and communication are design features of mobile health interventions and behavior change techniques [25] that can improve the effectiveness of the interventions. Morrison et al [28] also showed that personalization and communication are related to effective intervention outcomes. Therefore, these interactive design features can improve user participation and promote the effectiveness of interventions, and special attention should be given to them during the design of mobile health interventions.

In terms of presentation and navigation, users preferred a user friendly design, specifically, one that was easy to use and understand and was aesthetically pleasing, which was mainly addressed by the design of the interface aesthetic, navigation, and message presentation. A user friendly design is easy to use and understand, with features such as simple and convenient navigation, easy to understand language, and a supportive tone. Users often experience difficulties when using new technology [78], and an easy-to-use design can reduce that burden [79,80]. A study [41] showed the importance of a simple design when engaging users who were resistant to change. Compared to users who intend to change, those who were resistant to change were harder to engage in an intervention; however, this issue was alleviated with the help of a simple design [81].

Aesthetics also increases the friendliness of the design. A lack of aesthetics and text-heavy presentations made users feel bored, and users liked information presented with pictures and short texts. Additionally, a beautiful interface more easily attracts the user's attention. Studies have shown that a friendly mobile health intervention design can win the trust of users [82] and that credibility is a major concern for potential users [83]. Credibility needs to be established for the user to trust and use a mobile tool [84]. Tools that are developed in the future should provide evidence-based information, privacy policies, and password protection; they should also remove advertisements.

The 7 analytical themes from this study all appeared in different health topics, indicating that the design features proposed in this study were universal across the different health topics. Mobile apps and the website platforms shared 7 analytical themes, but text messages lacked reinforcement and interface aesthetics. The reason for this may be that the small number of articles involving text messages and the features designed to improve user engagement were not comprehensive and that text messages can be regarded as simple mobile apps, which cannot realize all of the functions of mobile apps; for example, the aesthetic interfaces design feature does not apply to text messages.

Mobile phone apps and website platforms are currently the most advanced mHealth technologies [85]. From this perspective, the design features of this study are also universal across different mobile technologies. Other simpler technologies can select a part of the design features in the checklist as a reference according to their own functions. Different mHealth technologies have different capabilities for presentation, navigation, and interaction, which may lead to different potential for improving user engagement. In the future, the relationship between different mHealth technologies and user engagement needs to be studied.

One study [86] about the design features of 100 smoking cessation apps on iTunes in 2016 showed that the existing smoking cessation apps performed well in terms of language but performed poorly in terms of presentation, navigation, and interaction. Our design feature checklist can help solve this problem. Researchers can use the checklist to guide the design process of mHealth interventions. The checklist can also be used to evaluate mHealth interventions that have been developed.

There is no unified terminology for design features. The same design feature has multiple different names in different studies. For example, communication can also be called *social support* and *social network*. This study summarized and translated the existing descriptions and described each design feature in a unified term to promote a standardized description of the design features.

Limitations

This study has the following limitations. First, we only selected studies that explained how a design feature is implemented in technology because some articles lacked detailed descriptions, making them impossible to analyze. However, this may mean that the design features we extracted do not comprehensively represent all research that has been performed. Second, most included articles were qualitative and mixed methods studies, which have considerable advantages given the exploratory nature of this research; however, compared with quantitative research, these methods provide less convincing evidence of the relationship between design features and user engagement. There may be sufficient studies available to explore design features that can improve engagement, and a more robust quantitative study design is needed to verify the association. Third, the mobile technology used in the studies in this review were mobile phone apps, website platforms, and text messages, which do not represent all technologies used in mHealth interventions. Other mobile tools, such as remote measurement technology and wearable devices, do not have as many functions as the aforementioned technologies included in this study. Subsets of the design features proposed in our research can also be applied to other forms of mobile technology, and relevant studies, especially regarding remote and wearable devices, are needed, considering their differences compared with the mobile phones, tablets and laptops used in this research. Fourth, the MMAT is currently the most applicable appraisal tool for a systematic mixed studies review, and it provides a detailed assessment of the quality of the included studies. The quality of the included studies varied, and 14 articles met 2 or 3 criteria of all 5 items, which means bias is present. However, the 2018 version MMAT does not provide quantification of a study's quality, and due to the heterogeneity of the included studies and the qualitative design of this systematic review, it is difficult to assess the weight of the different included studies or to obtain a more accurate understanding of the overall risk of bias.

Conclusions

This study summarized research results obtained in the past ten years to identify design features in mHealth interventions that improve user participation. We made a checklist that divided the design features of mHealth interventions into 7 different aspects with associated, clearly described implementations, which can not only be used as a reference during the mHealth development process but also as an evaluation tool for the design features of newly developed mHealth interventions. This checklist can be applied to mobile apps, website platforms, and text messages and can be applied to health topics such as unhealthy lifestyles, chronic diseases, and mental health problems. The study of the relationships between these design features and user engagement is in the exploratory stage but has great potential. We synthesized the results of currently available studies to promote better application of their results and to lay a foundation for subsequent confirmatory research.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (number 71573047).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the original studies.
[DOC File , 22 KB - jmir_v22i12e21687_app1.doc]

Multimedia Appendix 2
Quality assessment methodology.
[DOC File , 27 KB - jmir_v22i12e21687_app2.doc]

References

1. Källander K, Tibenderana JK, Akpogheneta OJ, Strachan DL, Hill Z, ten AAHA, et al. Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: a review. J Med Internet Res 2013;15(1):e17 [FREE Full text] [doi: 10.2196/jmir.2130] [Medline: 23353680]
2. Moessner M, Bauer S, Ozer F, Wolf M, Zimmer B, Kordy H. Cost-effectiveness of an internet-based aftercare intervention after inpatient treatment in a psychosomatic hospital. Psychother Res 2014;24(4):496-503. [doi: 10.1080/10503307.2013.845919] [Medline: 24188127]
3. Noar SM, Black HG, Pierce LB. Efficacy of computer technology-based HIV prevention interventions: a meta-analysis. AIDS 2009 Jan 2;23(1):107-115. [doi: 10.1097/QAD.0b013e32831c5500] [Medline: 19050392]

4. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To Prompt or Not to Prompt? A Microrandomized Trial of Time-Varying Push Notifications to Increase Proximal Engagement With a Mobile Health App. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
5. Ben-Zeev D, Scherer EA, Gottlieb JD, Rotondi AJ, Brunette MF, Achtyes ED, et al. mHealth for Schizophrenia: Patient Engagement With a Mobile Phone Intervention Following Hospital Discharge. JMIR Ment Health 2016;3(3):e34 [FREE Full text] [doi: 10.2196/mental.6348] [Medline: 27465803]
6. Muñoz RF, Aguilera A, Schueller SM, Leykin Y, Pérez-Stable EJ. From online randomized controlled trials to participant preference studies: morphing the San Francisco Stop Smoking site into a worldwide smoking cessation resource. J Med Internet Res 2012 Jun 27;14(3):e64 [FREE Full text] [doi: 10.2196/jmir.1852] [Medline: 22739225]
7. Newman MW, Lauterbach D, Munson SA, Resnick P, Morris ME. “It's not that I don't have problems, I'm just not putting them on Facebook”: Challenges and Opportunities in Using Online Social Networks for Health. 2011 Presented at: Proceedings of the 2011 ACM Conference on Computer Supported Cooperative Work; March 19-23; Hangzhou, China. [doi: 10.1145/1958824.1958876]
8. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: Complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016 Feb;59:299-307. [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]
9. Lustria MLA, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
10. Mouttapa M, Robertson TP, McEligot AJ, Weiss JW, Hoolihan L, Ora A, et al. The Personal Nutrition Planner: a 5-week, computer-tailored intervention for women. J Nutr Educ Behav 2011;43(3):165-172. [doi: 10.1016/j.jneb.2010.08.010] [Medline: 21550532]
11. BinDhim NF, McGeechan K, Trevena L. Who Uses Smoking Cessation Apps? A Feasibility Study Across Three Countries via Smartphones. JMIR Mhealth Uhealth 2014 Feb 06;2(1):e4. [doi: 10.2196/mhealth.2841]
12. Partridge SR, McGeechan K, Hebden L, Balestracci K, Wong AT, Denney-Wilson E, et al. Effectiveness of a mHealth Lifestyle Program With Telephone Support (TXT2BFiT) to Prevent Unhealthy Weight Gain in Young Adults: Randomized Controlled Trial. JMIR Mhealth Uhealth 2015;3(2):e66 [FREE Full text] [doi: 10.2196/mhealth.4530] [Medline: 26076688]
13. Recio-Rodriguez JI, Agudo Conde C, Calvo-Aponte MJ, Gonzalez-Viejo N, Fernandez-Alonso C, Mendizabal-Gallastegui N, EVIDENT Investigators. The Effectiveness of a Smartphone Application on Modifying the Intakes of Macro and Micronutrients in Primary Care: A Randomized Controlled Trial. The EVIDENT II Study. Nutrients 2018 Oct 10;10(10):1473 [FREE Full text] [doi: 10.3390/nu10101473] [Medline: 30309008]
14. Nguyen Thanh V, Guignard R, Lancrenon S, Bertrand C, Delva C, Berlin I, et al. Effectiveness of a Fully Automated Internet-Based Smoking Cessation Program: A Randomized Controlled Trial (STAMP). Nicotine Tob Res 2019 Jan 04;21(2):163-172. [doi: 10.1093/ntr/nty016] [Medline: 29370407]
15. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
16. Story A, Aldridge RW, Smith CM, Garber E, Hall J, Ferenando G, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 2019 Mar 23;393(10177):1216-1224 [FREE Full text] [doi: 10.1016/S0140-6736(18)32993-3] [Medline: 30799062]
17. Dillingham R, Ingersoll K, Flickinger TE, Waldman AL, Grabowski M, Laurence C, et al. PositiveLinks: A Mobile Health Intervention for Retention in HIV Care and Clinical Outcomes with 12-Month Follow-Up. AIDS Patient Care STDS 2018 Jun;32(6):241-250 [FREE Full text] [doi: 10.1089/apc.2017.0303] [Medline: 29851504]
18. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Dec;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
19. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Jun 05:116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
20. Kelders SM, Van GJEWC, Werkman A, Nijland N, Seydel ER. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
21. Augustson E, Cole-Lewis H, Sanders A, Schwarz M, Geng Y, Coa K, et al. Analysing user-reported data for enhancement of SmokefreeTXT: a national text message smoking cessation intervention. Tob Control 2017 Nov;26(6):683-689. [doi: 10.1136/tobaccocontrol-2016-052945] [Medline: 27852892]
22. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
23. Enwald HPK, Huotari MA. Preventing the obesity epidemic by second generation tailored health communication: an interdisciplinary review. J Med Internet Res 2010;12(2):e24 [FREE Full text] [doi: 10.2196/jmir.1409] [Medline: 20584698]

24. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and Nonusage Attrition With a Free Physical Activity Promotion Program: The Case of 10,000 Steps Australia. J Med Internet Res 2015;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
25. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
26. Schwarzer R. Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. Applied Psychology 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
27. Kok G, Schaalma H, Ruiter RAC, van EP, Brug J. Intervention mapping: protocol for applying health psychology theory to prevention programmes. J Health Psychol 2004 Jan;9(1):85-98. [doi: 10.1177/1359105304038379] [Medline: 14683571]
28. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
29. Ting Y. The Pitfalls of Mobile Devices in Learning: A Different View and Implications for Pedagogical Design. Journal of Educational Computing Research 2012 Jun 08;46(2):119-134. [doi: 10.2190/ec.46.2.a]
30. Lee S, Ha S, Widdows R. Consumer responses to high-technology products: Product attributes, cognition, and emotions. Journal of Business Research 2011 Nov;64(11):1195-1200. [doi: 10.1016/j.jbusres.2011.06.022]
31. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
32. Walther JB, Pingree S, Hawkins RP, Buller DB. Attributes of interactive online health information systems. J Med Internet Res 2005 Jul 1;7(3):e33 [FREE Full text] [doi: 10.2196/jmir.7.3.e33] [Medline: 15998624]
33. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
34. Crutzen R, de NJ, Brouwer W, Oenema A, Brug J, de Vries NK. A conceptual framework for understanding and improving adolescents' exposure to Internet-delivered interventions. Health Promot Int 2009 Sep;24(3):277-284. [doi: 10.1093/heapro/dap018] [Medline: 19515716]
35. Broderick J, Devine T, Langhans E, Lemerise AJ, Lier S, Harris L. Designing Health Literate Mobile Apps. In: NAM Perspectives. Washington, DC: Institute of Medicine of the National Academies; Jan 28, 2014.
36. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
37. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
38. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
39. Hong Q, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M. Mixed methods appraisal tool (MMAT), version 2018. Registration of Copyright 2018:1148552.
40. Su M, Chen W, Liu C, Jou H, Hsiao Y, Tsao L. [The Design Requirements for an E-Health Management Platform: Addressing the Needs of Adolescent Girls at High Risk of Metabolic Syndrome]. Hu Li Za Zhi 2015 Oct;62(5):51-60. [doi: 10.6224/JN62.5.51] [Medline: 26507627]
41. Crane D, Garnett C, Brown J, West R, Michie S. Factors Influencing Usability of a Smartphone App to Reduce Excessive Alcohol Consumption: Think Aloud and Interview Studies. Front Public Health 2017;5:39 [FREE Full text] [doi: 10.3389/fpubh.2017.00039] [Medline: 28421175]
42. Peters D, Deady M, Glozier N, Harvey S, Calvo RA. Worker Preferences for a Mental Health App Within Male-Dominated Industries: Participatory Study. JMIR Ment Health 2018 Apr 25;5(2):e30 [FREE Full text] [doi: 10.2196/mental.8999] [Medline: 29695371]
43. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
44. Lazard AJ, Pikowski J, Horrell L, Ross JC, Noar SM, Sutfin EL. Adolescents' and Young Adults' Aesthetics and Functionality Preferences for Online Tobacco Education. J Cancer Educ 2020 Apr;35(2):373-379. [doi: 10.1007/s13187-019-1475-4] [Medline: 30734263]
45. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Mirkovic J, Børøsund E, et al. Patients' Needs and Requirements for eHealth Pain Management Interventions: Qualitative Study. J Med Internet Res 2019 Apr 01;21(4):e13205 [FREE Full text] [doi: 10.2196/13205] [Medline: 30877780]

46. Milward J, Deluca P, Drummond C, Watson R, Dunne J, Kimergård A. Usability Testing of the BRANCH Smartphone App Designed to Reduce Harmful Drinking in Young Adults. JMIR Mhealth Uhealth 2017 Aug 08;5(8):e109 [FREE Full text] [doi: 10.2196/mhealth.7836] [Medline: 28790022]
47. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
48. Coyne I, Prizeman G, Sheehan A, Malone H, While AE. An e-health intervention to support the transition of young people with long-term illnesses to adult healthcare services: Design and early use. Patient Educ Couns 2016 Dec;99(9):1496-1504. [doi: 10.1016/j.pec.2016.06.005] [Medline: 27372524]
49. Peng W, Yuan S, Holtz BE. Exploring the Challenges and Opportunities of Health Mobile Apps for Individuals with Type 2 Diabetes Living in Rural Communities. Telemed J E Health 2016 Dec;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]
50. Gkatzidou V, Hone K, Sutcliffe L, Gibbs J, Sadiq ST, Szczepura A, et al. User interface design for mobile-based sexual health interventions for young people: design recommendations from a qualitative study on an online Chlamydia clinical care pathway. BMC Med Inform Decis Mak 2015 Aug 26;15:72 [FREE Full text] [doi: 10.1186/s12911-015-0197-8] [Medline: 26307056]
51. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. University Students’ Perspectives on Mindfulness and mHealth: A Qualitative Exploratory Study. American Journal of Health Education 2018 Sep 10;49(6):341-353. [doi: 10.1080/19325037.2018.1502701]
52. Phillips SM, Courneya KS, Welch WA, Gavin KL, Cottrell A, Nielsen A, et al. Breast cancer survivors' preferences for mHealth physical activity interventions: findings from a mixed methods study. J Cancer Surviv 2019 Apr;13(2):292-305. [doi: 10.1007/s11764-019-00751-3] [Medline: 30912011]
53. Herbeć A, Perski O, Shahab L, West R. Smokers' Views on Personal Carbon Monoxide Monitors, Associated Apps, and Their Use: An Interview and Think-Aloud Study. Int J Environ Res Public Health 2018 Feb 07;15(2):288 [FREE Full text] [doi: 10.3390/ijerph15020288] [Medline: 29414907]
54. Perski O, Baretta D, Blandford A, West R, Michie S. Engagement features judged by excessive drinkers as most important to include in smartphone applications for alcohol reduction: A mixed-methods study. Digit Health 2018;4:2055207618785841 [FREE Full text] [doi: 10.1177/2055207618785841] [Medline: 31463077]
55. Utrankar A, Mayo-Gamble TL, Allen W, Novak L, Kassim AA, Bonnet K, et al. Technology use and preferences to support clinical practice guideline awareness and adherence in individuals with sickle cell disease. J Am Med Inform Assoc 2018 May 07:976-988. [doi: 10.1093/jamia/ocy036] [Medline: 29741695]
56. Grunberg PH, Dennis C, Da Costa D, Zelkowitz P. Infertility patients' need and preferences for online peer support. Reprod Biomed Soc Online 2018 Aug;6:80-89 [FREE Full text] [doi: 10.1016/j.rbms.2018.10.016] [Medline: 30547107]
57. Willoughby JF, Muldrow A. SMS for sexual health: A comparison of service types and recommendations for sexual health text message service providers. Health Education Journal 2016 Sep 27;76(2):231-243. [doi: 10.1177/0017896916661373]
58. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User Preferences and Design Recommendations for an mHealth App to Promote Cystic Fibrosis Self-Management. JMIR Mhealth Uhealth 2014;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
59. Hartzler AL, BlueSpruce J, Catz SL, McClure JB. Prioritizing the mHealth Design Space: A Mixed-Methods Analysis of Smokers' Perspectives. JMIR Mhealth Uhealth 2016 Aug 05;4(3):e95 [FREE Full text] [doi: 10.2196/mhealth.5742] [Medline: 27496593]
60. Peters D, Davis S, Calvo RA, Sawyer SM, Smith L, Foster JM. Young People's Preferences for an Asthma Self-Management App Highlight Psychological Needs: A Participatory Study. J Med Internet Res 2017 Apr 11;19(4):e113 [FREE Full text] [doi: 10.2196/jmir.6994] [Medline: 28400353]
61. McClure JB, Heffner J, Hohl S, Klasnja P, Catz SL. Design Considerations for mHealth Programs Targeting Smokers Not Yet Ready to Quit: Results of a Sequential Mixed-Methods Study. JMIR Mhealth Uhealth 2017 Mar 10;5(3):e31 [FREE Full text] [doi: 10.2196/mhealth.6845] [Medline: 28283465]
62. Evans C, Turner K, Suggs LS, Occa A, Juma A, Blake H. Developing a mHealth intervention to promote uptake of HIV testing among African communities in the UK: a qualitative study. BMC Public Health 2016 Dec 28;16:656 [FREE Full text] [doi: 10.1186/s12889-016-3278-4] [Medline: 27465586]
63. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
64. Carolan S, de Visser RO. Employees' Perspectives on the Facilitators and Barriers to Engaging With Digital Mental Health Interventions in the Workplace: Qualitative Study. JMIR Ment Health 2018 Jan 19;5(1):e8 [FREE Full text] [doi: 10.2196/mental.9146] [Medline: 29351900]
65. Thornton LK, Kay-Lambkin FJ. Specific features of current and emerging mobile health apps: user views among people with and without mental health problems. Mhealth 2018;4:56 [FREE Full text] [doi: 10.21037/mhealth.2018.11.04] [Medline: 30701174]

66. Saberi P, Siedle-Khan R, Sheon N, Lightfoot M. The Use of Mobile Health Applications Among Youth and Young Adults Living with HIV: Focus Group Findings. AIDS Patient Care STDS 2016 Jun;30(6):254-260 [FREE Full text] [doi: 10.1089/apc.2016.0044] [Medline: 27214751]
67. Aji M, Gordon C, Peters D, Bartlett D, Calvo RA, Naqshbandi K, et al. Exploring User Needs and Preferences for Mobile Apps for Sleep Disturbance: Mixed Methods Study. JMIR Ment Health 2019 May 24;6(5):e13895 [FREE Full text] [doi: 10.2196/13895] [Medline: 31127714]
68. Geuens J, Geurts L, Swinnen TW, Westhovens R, Vanden Abeele V. Mobile Health Features Supporting Self-Management Behavior in Patients With Chronic Arthritis: Mixed-Methods Approach on Patient Preferences. JMIR Mhealth Uhealth 2019 Mar 25;7(3):e12535 [FREE Full text] [doi: 10.2196/12535] [Medline: 30907737]
69. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a Mobile HIV Prevention App for Men Who Have Sex With Men. JMIR Mhealth Uhealth 2014;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
70. McClure JB, Shortreed SM, Bogart A, Derry H, Riggs K, St John J, et al. The Effect of Program Design on Engagement With an Internet-Based Smoking Intervention: Randomized Factorial Trial. J Med Internet Res 2013 Mar 25;15(3):e69. [doi: 10.2196/jmir.2508]
71. Wright CJC, Dietze PM, Lim MSC. Beyond Basic Feedback in Mobile Brief Interventions: Designing SMS Message Content for Delivery to Young Adults During Risky Drinking Events. JMIR Mhealth Uhealth 2017 Jun 20;5(6):e79 [FREE Full text] [doi: 10.2196/mhealth.6497] [Medline: 28634153]
72. Fylan F, Caveney L, Cartwright A, Fylan B. Making it work for me: beliefs about making a personal health record relevant and useable. BMC Health Serv Res 2018 Jun 14;18(1):445 [FREE Full text] [doi: 10.1186/s12913-018-3254-z] [Medline: 29898730]
73. Gatwood J, Shuvo S, Ross A, Riordan C, Smith P, Gutierrez ML, et al. The Management of Diabetes in Everyday Life (MODEL) program: development of a tailored text message intervention to improve diabetes self-care activities among underserved African-American adults. Transl Behav Med 2020 Feb 03;10(1):204-212. [doi: 10.1093/tbm/ibz024] [Medline: 30794316]
74. Zhao Y, Zhu X, Pérez AE, Zhang W, Shi A, Zhang Z, et al. MHealth approach to promote Oral HIV self-testing among men who have sex with men in China: a qualitative description. BMC Public Health 2018 Sep 27;18(1):1146. [doi: 10.1186/s12889-018-6046-9]
75. Tîrnăucă C, Duque R, Montaña JL. User Interaction Modeling and Profile Extraction in Interactive Systems: A Groupware Application Case Study. Sensors (Basel) 2017 Jul 20;17(7):1669 [FREE Full text] [doi: 10.3390/s17071669] [Medline: 28726762]
76. Appleton KM, Bray J, Price S, Liebchen G, Jiang N, Mavridis I, et al. A Mobile Phone App for the Provision of Personalized Food-Based Information in an Eating-Out Situation: Development and Initial Evaluation. JMIR Form Res 2019 Nov 04;3(4):e12966 [FREE Full text] [doi: 10.2196/12966] [Medline: 31682575]
77. Digital Health Center of Excellence. US Food and Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health [accessed 2020-03-26]
78. Lyles CR, Harris LT, Le T, Flowers J, Tufano J, Britt D, et al. Qualitative evaluation of a mobile phone and web-based collaborative care intervention for patients with type 2 diabetes. Diabetes Technol Ther 2011 May;13(5):563-569. [doi: 10.1089/dia.2010.0200] [Medline: 21406018]
79. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
80. Krug S. Don't Make Me Think!: A Common Sense Approach to Web Usability. Indianapolis, United States: Que Corp; 2000.
81. Lambert MJ, Barley DE. Research summary on the therapeutic relationship and psychotherapy outcome. Psychotherapy: Theory, Research, Practice, Training 2001;38(4):357-361. [doi: 10.1037/0033-3204.38.4.357]
82. Lowry PB, Wilson DW, Haig WL. A Picture is Worth a Thousand Words: Source Credibility Theory Applied to Logo and Website Design for Heightened Credibility and Consumer Trust. International Journal of Human-Computer Interaction 2013 Nov 27;30(1):63-93. [doi: 10.1080/10447318.2013.839899]
83. Lorimer K, McDaid L. Young men's views toward the barriers and facilitators of Internet-based Chlamydia trachomatis screening: qualitative study. J Med Internet Res 2013;15(12):e265 [FREE Full text] [doi: 10.2196/jmir.2628] [Medline: 24300158]
84. Fogg BJ. Prominence-interpretation theory: Explaining how people assess credibility online. CHI'03 extended abstracts on human factors in computing systems 2003:722-723. [doi: 10.1145/765891.765951]
85. Bhavnani SP, Narula J, Sengupta PP. Mobile technology and the digitization of healthcare. Eur Heart J 2016 Feb 11;37(18):1428-1438. [doi: 10.1093/eurheartj/ehv770]
86. Paige SR, Alber JM, Stellefson ML, Krieger JL. Missing the mark for patient engagement: mHealth literacy strategies and behavior change processes in smoking cessation apps. Patient Educ Couns 2018 May;101(5):951-955. [doi: 10.1016/j.pec.2017.11.006] [Medline: 29153592]

Abbreviations

mHealth: mobile health
MMAT: Mixed Methods Appraisal Tool

Edited by G Eysenbach; submitted 22.06.20; peer-reviewed by S Six, M Rauws; comments to author 10.07.20; revised version received 30.09.20; accepted 24.10.20; published 09.12.20.

Please cite as:
Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H
Design Features for Improving Mobile Health Intervention User Engagement: Systematic Review and Thematic Analysis
J Med Internet Res 2020;22(12):e21687
URL: https://www.jmir.org/2020/12/e21687
doi:10.2196/21687
PMID:33295292

Review

Intimate Partner Violence and Electronic Health Interventions: Systematic Review and Meta-Analysis of Randomized Trials

Ditte S Linde[1,2,3,4], BSc, MSc, PhD; Aleksandra Bakiewicz[1,2], BA, MA; Anne Katrine Normann[1], MD; Nina Beck Hansen[5], BSc, MSc, PhD; Andreas Lundh[1,4,6,7], PhD, MD; Vibeke Rasch[1,2], DMSci, MD, PhD

[1]Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[2]Department of Gynaecology & Obstetrics, Odense University Hospital, Odense, Denmark
[3]Department of Public Health, University of Southern Denmark, Esbjerg, Denmark
[4]Odense Patient Data Explorative Network, Odense University Hospital, Odense, Denmark
[5]Department of Psychology, University of Southern Denmark, Odense, Denmark
[6]Centre for Evidence-Based Medicine Odense, Odense University Hospital, Odense, Denmark
[7]Department of Infectious Diseases, Hvidovre Hospital, Copenhagen, Denmark

Corresponding Author:
Ditte S Linde, BSc, MSc, PhD
Department of Clinical Research
University of Southern Denmark
Kløvervænget 10, 10
Odense
Denmark
Phone: 45 61666564
Email: dsondergaard@health.sdu.dk

Abstract

Background: Intimate partner violence (IPV) is a major public health concern. eHealth interventions may reduce exposure to violence and health-related consequences as the technology provides a safe and flexible space for the target population. However, the evidence is unclear.

Objective: The goal of the review is to examine the effect of eHealth interventions compared with standard care on reducing IPV, depression, and posttraumatic stress disorder (PTSD) among women exposed to IPV.

Methods: We searched EMBASE, MEDLINE, Cochrane Central Register of Controlled Trials, PsycInfo, Scopus, Global Health Library, ClinicalTrials.gov, and International Clinical Trials Registry Platform for published and unpublished trials from inception until April 2019. Trials with an eHealth intervention targeting women exposed to violence were included. We assessed risk of bias using the Cochrane Risk of Bias Tool. Trials that reported effect estimates on overall IPV; physical, sexual, and psychological violence; depression; or posttraumatic stress disorder were included in meta-analyses.

Results: A total of 14 trials were included in the review; 8 published trials, 3 unpublished trials and 3 ongoing trials. Of the 8 published trials, 2 were judged as overall low risk of bias trials. The trials reported 23 types of outcomes, and 7 of the trials had outcomes that were eligible for meta-analyses. Our pooled analyses found no effect of eHealth interventions on any of our prespecified outcomes: overall IPV (SMD –0.01; 95% CI –0.11 to 0.08; I^2=0%; 5 trials, 1668 women); physical violence (SMD 0.01; 95% CI –0.22 to 0.24; I^2=58%; 4 trials, 1128 women); psychological violence (SMD 0.07; 95% CI –0.12 to 0.25; I^2=40%; 4 trials, 1129 women); sexual violence (MD 0.36; 95% CI –0.18 to 0.91; I^2=0%; 2 trials, 1029 women); depression (SMD –0.13; 95% CI –0.37 to 0.11; I^2=78%; 5 trials, 1600 women); and PTSD (MD –0.11; 95% CI –1.04 to 0.82; I^2=0%; 5 trials, 1267 women).

Conclusions: There is no evidence from randomized trials of a beneficial effect of eHealth interventions on IPV. More high-quality trials are needed, and we recommend harmonizing outcome reporting in IPV trials by establishing core outcome sets.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019130124; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=130124

(J Med Internet Res 2020;22(12):e22361) doi:10.2196/22361

KEYWORDS

eHealth; randomized trials; intimate partner violence; domestic violence; abuse; depression; PTSD

Introduction

Background

Intimate partner violence (IPV) is defined as “a behavior by an intimate partner or ex-partner that causes physical, sexual, or psychological harm, including physical aggression, sexual coercion, psychological abuse, and controlling behaviors” [1]. It is also known as domestic abuse, domestic violence, or battering, and it is a major public health issue and a violation of human rights [2,3]. IPV can affect both men and women, yet most survivors are women [4,5]. Research on the prevalence of male survivors of IPV is scarce, and to the best of our knowledge, there is currently no global estimate on the magnitude of problem. However, a 2015 national survey from the United States reported that 11% of American men experienced some form of IPV during their lifetime [6]. Globally, approximately 1 in 3 women will experience physical or sexual violence from their partner during their lifetime. However, there are regional differences with the highest prevalence being found in Southeast Asia, the Eastern Mediterranean region, and Africa (around 37%) while the lowest prevalence is found in high-income countries (around 23%) [7].

IPV can have a number of immediate and long-term health consequences including physical injury, depression, anxiety, posttraumatic stress disorder (PTSD), suicidality, and substance abuse as well as gastrointestinal and gynecologic problems [7,8]. The worst cases can lead to homicide [9]. Further, the fetuses or children of the IPV survivors may be indirectly exposed to IPV, which can result in induced abortion, preterm birth, low birth weight, and infant mortality as well as developmental and behavioral problems later in life [10]. Often people experiencing IPV do not report the violence or delay seeking counseling due to a number of barriers, including stigma, embarrassment, and fear of the perpetrator [2].

eHealth is defined as the use of information and communication technologies for health [11]. It is a diverse concept that encompasses the subareas mobile health (mHealth) and telehealth [12]. It has been hypothesized that eHealth interventions have potential to reduce IPV exposure and its health-related consequences as the technology provides a safe and flexible space for the target population compared with traditional face-to-face approaches [13]. However, evidence of the effect of eHealth on IPV is unclear. Two Cochrane reviews from 2014 and 2015 assessed interventions for prevention and reduction of IPV among pregnant women [14] and women in general [15]. Some eHealth interventions were included in these reviews and showed mixed results [16-18]. New trials have since been published, and to our knowledge there is no systematic review specifically addressing eHealth interventions and their effect on reducing IPV and IPV-related health consequences.

Objectives

The goal of the review is to estimate the effect of eHealth interventions compared with standard care on reducing overall IPV (physical, sexual, or psychological violence), type-specific IPV, depression, and PTSD among women exposed to IPV.

Methods

Protocol and Registration

The protocol was registered at the International Prospective Register for Systematic Reviews (PROSPERO) prior to study conduct [CRD42019130124] (registration date: April 15, 2019) [19]. The review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 checklist [20] (Multimedia Appendix 1).

Eligibility Criteria

We included published and unpublished randomized controlled trials, including pilot trials, in any language and setting. Further, we included trials of women exposed to any type of IPV by a current or former partner at any point in life. All types of eHealth interventions (eg, videos, text messages or social media interventions) were included and eHealth interventions had to be compared with standard of care, placebo-like interventions (eg, online counseling on another health issue than IPV), other eHealth interventions, or another type of interventions (eg, face-to-face counseling). We excluded trials of survivors of other forms of violence (eg, dating violence or gang violence), trials restricted to survivors of IPV with substance problems or sexual minorities, and trials targeting both men and women if separate data for women were not available.

Information Sources and Search Strategy

We searched EMBASE, MEDLINE, PsycInfo, Scopus, Cochrane Central Register of Controlled Trials, and Global Health Library for trials from inception up to April 2019 (Multimedia Appendix 2). The search strategy was developed in collaboration with an experienced research librarian. In addition, we searched reference lists of included trials, the International Clinical Trials Registry Platform, and ClinicalTrials.gov in June 2019 for unpublished or ongoing trials.

Outcomes

Our primary outcome was overall IPV (physical and/or sexual and/or psychological violence). Our secondary outcomes were type-specific IPV (ie, physical violence, psychological violence, and sexual violence), depression, and PTSD.

Study Selection

After removing duplicates, two authors (AB, AKNN) screened titles and abstracts for obvious exclusion and assessed full-text papers using the web-based systematic review production tool Covidence [21]. Disagreements were resolved through discussion, and there was no need for involvement of an arbiter.

Data Extraction

Two authors (AB, AKNN) identified relevant outcomes, and one author (AB) extracted data verbatim into a standardized Excel (Microsoft Corp) template. One author (DSL) extracted outcome data for meta-analysis and verified the other data. Extracted data included first author, publication year, title, journal name, registry record ID, length of study, country, setting, objective, eligibility criteria, number of participants, number of males and females, mean age, description of interventions, primary and secondary outcomes, and funding source. Corresponding authors were contacted for unpublished data.

Risk of Bias Assessment

Two authors (AB, AKNN) independently assessed published trials for risk of bias using the Cochrane Risk of Bias Tool [22]. The following domains were assessed: sequence generation and allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessors (detection bias), incomplete outcome data (attrition bias), and selective outcome reporting (reporting bias). Domains were assessed as having low risk, high risk, or unclear risk of bias. Trials were judged as overall low risk of bias if they had low risk of selection bias, detection bias, and reporting bias. All other trials were judged as having high risk of bias. Disagreements were resolved through discussion. In case of disagreements, a third coauthor (DSL) made a final decision.

Data Analysis

For our descriptive analysis of study outcomes and outcome measurements scales, we constructed a multiple outcome matrix using the methodology developed by Mayo-Wilson and colleagues [23]. Meta-analyses were conducted on reduction of IPV (overall or physical, sexual, or psychological violence), PTSD, and depression. Meta-analyses were done using RevMan 5.3 (Cochrane). We planned to use both continuous and dichotomous outcome data, but no trials reported dichotomous outcome data. As we expected trials to be heterogeneous in terms of methodology, types of populations, and interventions, we used random effects models and the inverse-variance method. If trials reported continuous data using the same outcome measure (ie, similar scale), we analyzed data using mean difference, and if trials used different scales, we analyzed data using standardized mean difference and calculated corresponding 95% confidence intervals. We assessed statistical heterogeneity by using I^2. If trials had several time points for follow-up, we used the latest time point in our analyses. We conducted subgroup analyses comparing overall low risk of bias trials with high risk of bias trials, type of eHealth intervention, and type of scale for our primary outcome.

Results

Summary

We identified 1683 unique records, and excluded 1589 records after screening titles and abstracts (Figure 1). Of the 94 records reviewed in full text, 83 were excluded, leaving 11 trials for inclusion [8,24-36]. Three additional trials were included from searching other sources, leading to the inclusion of 14 trials in the review. Of the 14 trials, 8 were finished and published [8,24-31], 3 were finished but unpublished [32,35,37], and 3 were ongoing [33,34,36] (Table 1). Nine corresponding authors were contacted for clarification of data or unpublished data [8,28,29,32-35,37,38]. Seven authors replied [28,29,32-35,37], and 2 provided unpublished data [29,35] in the form of a different standard deviation, which was used in the meta-analysis [29], and tabulated data for a finished trial in the form of a draft manuscript [35]. However, as we were unable to resolve queries concerning the data, we decided not to include the data in our review.

The 8 trials were published from 2002 to 2019 enrolling 2147 women in total (median 202 participants per trial; Table 1). In the 6 trials that were either unpublished or ongoing, 3966 women were planned to be enrolled (median 450 participants per trial). The published trials were conducted in the United States (n=6), Australia (n=1), and New Zealand (n=1) and, except for 1 study that targeted couples [31], solely included women. The mean age of the participants ranged from 27.6 to 40.0 years, and follow-up varied from 1.5 to 12 months. Recruitment strategies varied across trials from general advertisements on television or online spaces to more specific advertisement in family court waiting areas and health clinics. All trials were 2-arm except for one 3-arm trial [28]. Three trials compared an online safety decision aid with a control website or standard safety planning [8,24,25], 1 trial compared online education on IPV with online popular TV shows [27], 2 trials assessed telephone support compared with standard care [29,30], 1 trial compared email modules to placebo email modules [31], and the 3-armed trial compared email modules to standard care or face-to-face modules [28]. Types of outcomes and how they were measured differed greatly across the 14 trials; 23 (median 4; interquartile range 3.75) types of outcomes and 49 outcome measurements were reported (Figure 2). For example, 7 different scales were used to measure self-efficacy, 5 different scales were used to measure overall IPV, and 4 different scales were used to measure depression.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart.

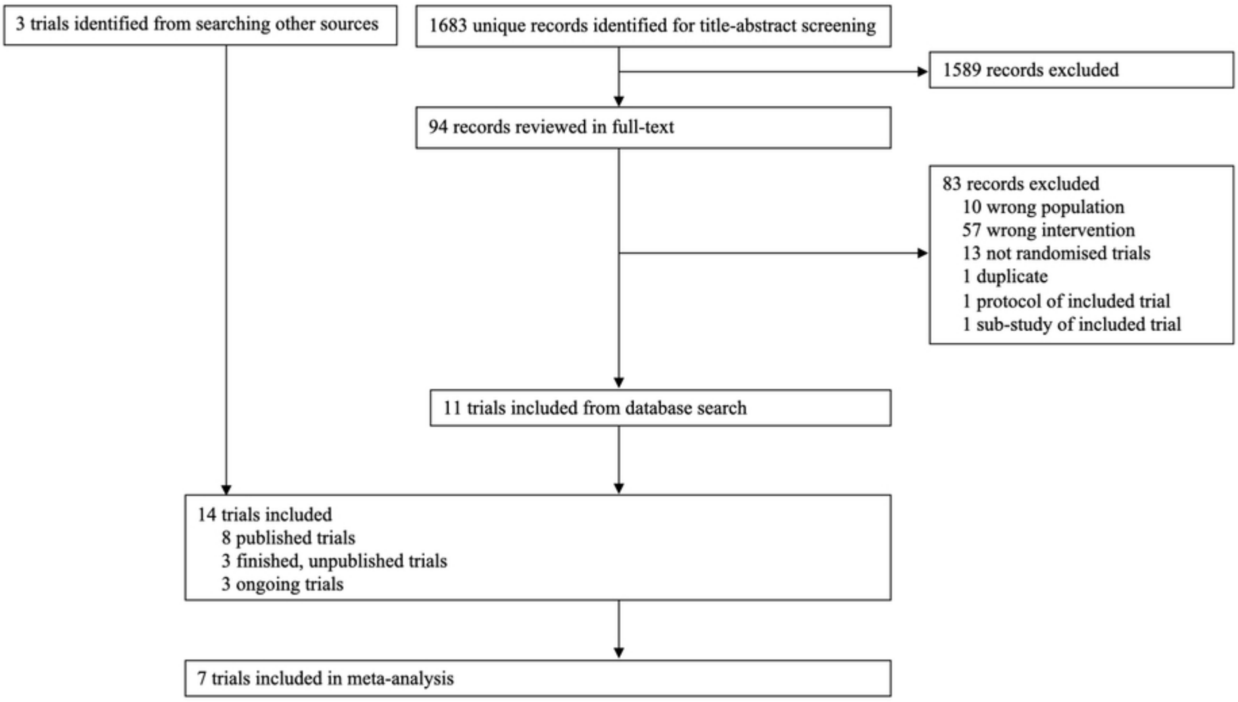


Table 1. Characteristics of trials included in the review.

Trial type	Country	Trial size (n)	Women, %	Age in years, mean	Follow-up (months)	Recruitment	Intervention[a]	Comparator	Primary outcome measure (scale)
Finished, published									
Hegarty [24]	Australia	422	100	33.7	12	Online advertisement; compensation for time up to Aus $150 (US $110)	Online safety decision aid	Control website	Self-efficacy (GSE[b])
Koziol-McLain [25]	New Zealand	412	100	29.0 (median)	12	TV advertisements and flyers at health clinics	Online safety decision aid	Control website	Depression (CES-D[c])
Zlotnick [27]	US	53	100	27.6	3	Pregnant women seeking mental health care who screened positive for IPV[d]	Online education on IPV	Online popular TV shows	Satisfaction with intervention (CSQ-8-R[e])
Glass [8]	US	721	100	33.4	12	Online advertisement, flyers at health clinics and public toilets	Online safety decision aid	Control website	IPV (SVAWS[f], WEB[g])
Constantino [28]	US	32	100	40	1.5	Family court waiting areas, legal services, women's shelters	email modules with IPV support (arm 1) or face-to-face modules with IPV support (arm 2)	Standard care	Anxiety (PROMIS[h])
Stevens [29]	US	253	100	29.2	6	Women at pediatric emergency departments who screened positive for IPV	Telephone support	Standard care	IPV (CAS[i]; WEB)
Braithwaite [31]	US	104	50	32.4	12	Online, posters, and newspaper advertisements	Emails, modules with relationship communication skills, and problem-solving training	Placebo emails; modules with information about depression, anxiety, and healthy relationships	Physical and psychological violence (CTS[j])
McFarlane [30]	US	150	100	30.3	6	Family violence unit	Telephone support	Standard care	Safety behavior
Finished, unpublished									
Clark [35] (NCT02942433, retrospectively registered)	Nepal	1440[k] (36 clusters)	50	—	18	Women participating in survey at development centers	Weekly radio drama, SMS[l], phone calls, and discussion groups	Weekly radio drama, SMS	Physical and/or sexual violence (unspecified)
Ford-Gilboe [32]	Canada	450	100	—	12	Advertisements in various online spaces	Online safety decision aid	General online safety information	Depression (CES-D)

Trial type	Country	Trial size (n)	Women, %	Age in years, mean	Follow-up (months)	Recruitment	Intervention[a]	Comparator	Primary outcome measure (scale)
PACTR201804003321122 [37]	Kenya	450	100	—	3	Study centers in Nairobi settlement	App with safety decision aid	Standard care	Sexual and reproductive coercion (unspecified)
Ongoing									
Henriksen [34] (NCT03397277)	Norway	525	100	—	3	Women attending antenatal clinics who screened positive for IPV	Safety decision aid video	Control video	Safety behavior (McFarlane's list)
Sabri [33] (NCT03265847)	US	1250	100	—	12	Written/verbal invitation to indigenous, immigrant, and refugee women; invitations sent through list servers, emails, and snowballing	Online and app safety decision aid	Control website	Physical violence (CTS-2)
NTR7313 [36]	Netherlands	198	100	—	6	Women self-identifying as IPV survivors through questions and registration online for SAFE (eHealth intervention)	Online safety decision aid	Not reported	Self-efficacy (GSE)

[a] Study with 2 intervention arms is specified by arm 1 and arm 2. Other trials had 1 intervention arm that could consist of multiple elements.

[b] GSE: General Self-Efficacy Scale.

[c] CESD: Center for Epidemiologic Studies Depression Scale.

[d] IPV: intimate partner violence.

[e] CSQ-8-R: Client Satisfaction Questionnaire, Revised–8 item.

[f] SVAWS: Severity of Violence Against Women Scale.

[g] WEB: Women's Experience With Battering Scale.

[h] PROMIS: Patient-Reported Outcomes Measurement Information System.

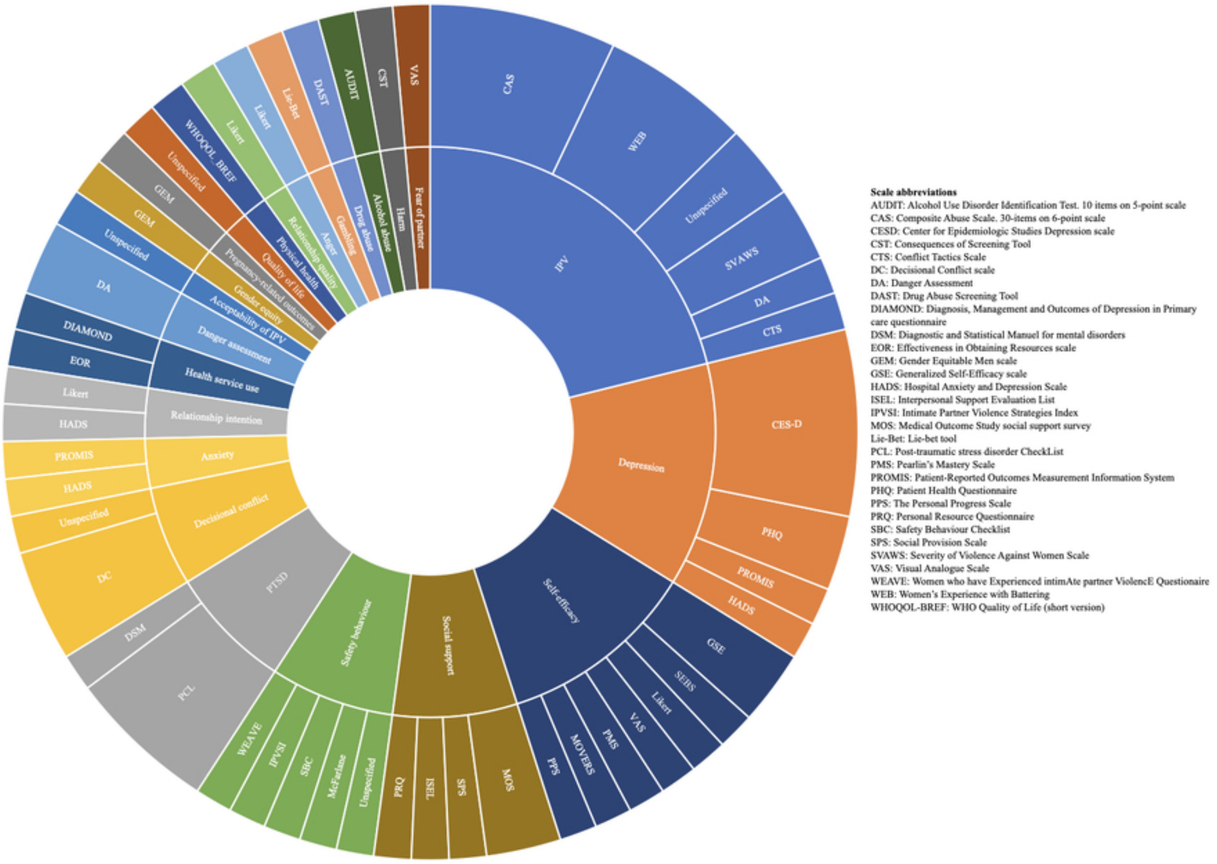
[i] CAS: Composite Abuse Scale.

[j] CTS: Conflict Tactics Scale.

[k] Unpublished data reported by corresponding author.

[l] SMS: short message service (text messaging).

Figure 2. Outcomes and outcome measurement scales in trials with eHealth interventions and intimate partner violence.

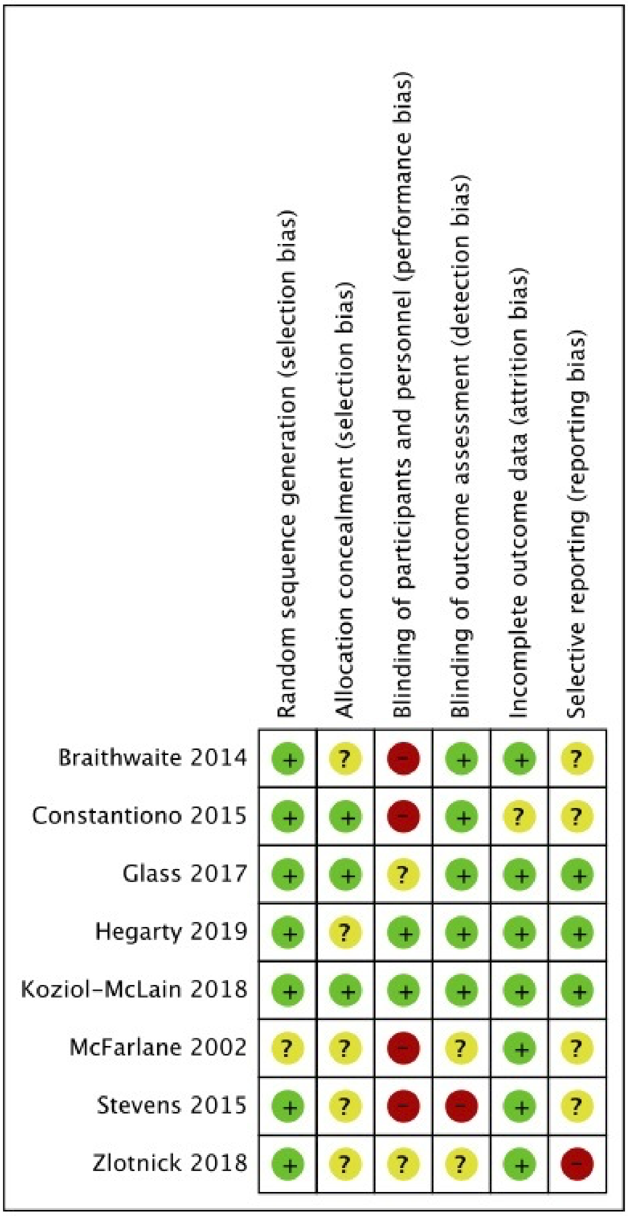


Scale abbreviations
AUDIT: Alcohol Use Disorder Identification Test. 10 items on 5-point scale
CAS: Composite Abuse Scale. 30-items on 6-point scale
CESD: Center for Epidemiologic Studies Depression scale
CST: Consequences of Screening Tool
CTS: Conflict Tactics Scale
DC: Decisional Conflict scale
DA: Danger Assessment
DAST: Drug Abuse Screening Tool
DIAMOND: Diagnosis, Management and Outcomes of Depression in Primary care questionnaire
DSM: Diagnostic and Statistical Manuel for mental disorders
EOR: Effectiveness in Obtaining Resources scale
GEM: Gender Equitable Men scale
GSE: Generalized Self-Efficacy scale
HADS: Hospital Anxiety and Depression Scale
ISEL: Interpersonal Support Evaluation List
IPVSI: Intimate Partner Violence Strategies Index
MOS: Medical Outcome Study social support survey
Lie-Bet: Lie-bet tool
PCL: Post-traumatic stress disorder CheckList
PMS: Pearlin's Mastery Scale
PROMIS: Patient-Reported Outcomes Measurement Information System
PHQ: Patient Health Questionnaire
PPS: The Personal Progress Scale
PRQ: Personal Resource Questionnaire
SBC: Safety Behaviour Checklist
SPS: Social Provision Scale
SVAWS: Severity of Violence Against Women Scale
VAS: Visual Analogue Scale
WEAVE: Women who have Experienced intimAte partner ViolencE Questionaire
WEB: Women's Experience with Battering
WHOQOL-BREF: WHO Quality of Life (short version)

Risk of Bias

Of the 8 published trials, 2 were assessed as overall in low risk of bias [24,25]. In 5 trials, allocation concealment was judged to be unclear, and lack of blinding or unclear description of blinding of personnel resulted in only 2 of the 8 trials being judged as in low risk of performance bias. Further, 4 trials did not have a record in a trial registry, which led to a judgment of unclear risk of reporting bias, and 1 trial had outcomes in the registry not reported in the trial publication, which led to a judgment of high risk of reporting bias (Figure 3, Multimedia Appendix 3).

Figure 3. Risk of bias assessment.

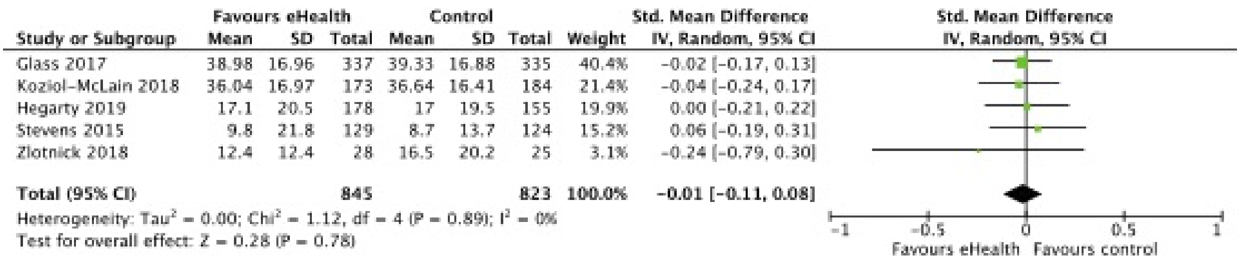


	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)
Braithwaite 2014	+	?	−	+	+	?
Constantiono 2015	+	+	−	+	?	?
Glass 2017	+	+	?	+	+	+
Hegarty 2019	+	?	+	+	+	+
Koziol-McLain 2018	+	+	+	+	+	+
McFarlane 2002	?	?	−	?	+	?
Stevens 2015	+	?	−	−	+	?
Zlotnick 2018	+	?	?	?	+	−

Meta-Analysis

Seven published trials were eligible for meta-analysis as they had outcomes on either overall IPV, type-specific IPV, depression, or PTSD [8,24,25,27-29,31]. Five trials (1668 participants) reported data on our primary outcome, overall IPV, and we found no difference in effect of eHealth compared with no eHealth interventions (standardized mean difference [SMD] –0.01; 95% CI –0.11 to 0.08; I^2=0% [Figure 4]). Four trials reported data on physical violence (1128 participants) and psychological violence (1129 participants) [8,25,27,31], and we found no difference in effect of eHealth interventions compared with no eHealth interventions (SMD [physical] 0.01; 95% CI –0.22 to 0.24; I^2=58% [Figure 5]; SMD [psychological] 0.07; 95% CI –0.12 to 0.25; I^2=40% [Figure 6]). Two trials (1029 participants) reported data on sexual violence [8,25], and we found no effect of eHealth interventions compared with no eHealth interventions (mean difference [MD] 0.36; 95% CI –0.18 to 0.91; I^2=0% [Severity of Violence Against Women Scale; Figure 7]).

Figure 4. Effect of eHealth versus no eHealth on overall intimate partner violence.

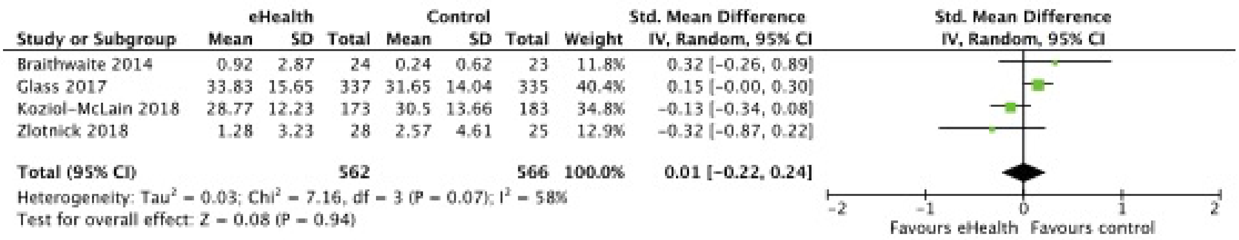


Study or Subgroup	Favours eHealth Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Glass 2017	38.98	16.96	337	39.33	16.88	335	40.4%	-0.02 [-0.17, 0.13]
Koziol-McLain 2018	36.04	16.97	173	36.64	16.41	184	21.4%	-0.04 [-0.24, 0.17]
Hegarty 2019	17.1	20.5	178	17	19.5	155	19.9%	0.00 [-0.21, 0.22]
Stevens 2015	9.8	21.8	129	8.7	13.7	124	15.2%	0.06 [-0.19, 0.31]
Zlotnick 2018	12.4	12.4	28	16.5	20.2	25	3.1%	-0.24 [-0.79, 0.30]
Total (95% CI)			**845**			**823**	**100.0%**	**-0.01 [-0.11, 0.08]**

Heterogeneity: Tau^2 = 0.00; Chi^2 = 1.12, df = 4 (P = 0.89); I^2 = 0%

Test for overall effect: Z = 0.28 (P = 0.78)

Figure 5. Effect of eHealth versus no eHealth on physical violence.

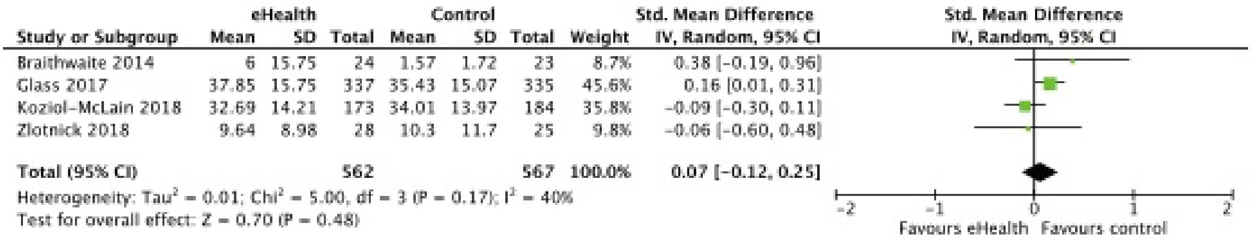


Study or Subgroup	eHealth Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Braithwaite 2014	0.92	2.87	24	0.24	0.62	23	11.8%	0.32 [-0.26, 0.89]
Glass 2017	33.83	15.65	337	31.65	14.04	335	40.4%	0.15 [-0.00, 0.30]
Koziol-McLain 2018	28.77	12.23	173	30.5	13.66	183	34.8%	-0.13 [-0.34, 0.08]
Zlotnick 2018	1.28	3.23	28	2.57	4.61	25	12.9%	-0.32 [-0.87, 0.22]
Total (95% CI)			**562**			**566**	**100.0%**	**0.01 [-0.22, 0.24]**

Heterogeneity: Tau^2 = 0.03; Chi^2 = 7.16, df = 3 (P = 0.07); I^2 = 58%

Test for overall effect: Z = 0.08 (P = 0.94)

Figure 6. Effect of eHealth versus no eHealth on psychological violence.

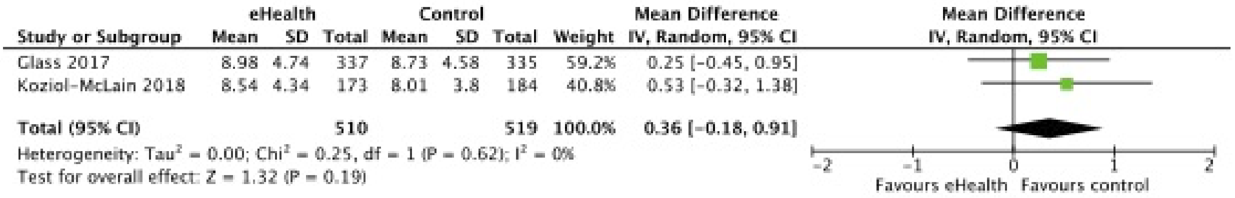


Study or Subgroup	eHealth Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Braithwaite 2014	6	15.75	24	1.57	1.72	23	8.7%	0.38 [-0.19, 0.96]
Glass 2017	37.85	15.75	337	35.43	15.07	335	45.6%	0.16 [0.01, 0.31]
Koziol-McLain 2018	32.69	14.21	173	34.01	13.97	184	35.8%	-0.09 [-0.30, 0.11]
Zlotnick 2018	9.64	8.98	28	10.3	11.7	25	9.8%	-0.06 [-0.60, 0.48]
Total (95% CI)			**562**			**567**	**100.0%**	**0.07 [-0.12, 0.25]**

Heterogeneity: Tau^2 = 0.01; Chi^2 = 5.00, df = 3 (P = 0.17); I^2 = 40%

Test for overall effect: Z = 0.70 (P = 0.48)

Figure 7. Effect of eHealth versus no eHealth on sexual violence.



Study or Subgroup	eHealth Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Glass 2017	8.98	4.74	337	8.73	4.58	335	59.2%	0.25 [-0.45, 0.95]
Koziol-McLain 2018	8.54	4.34	173	8.01	3.8	184	40.8%	0.53 [-0.32, 1.38]
Total (95% CI)			**510**			**519**	**100.0%**	**0.36 [-0.18, 0.91]**

Heterogeneity: Tau^2 = 0.00; Chi^2 = 0.25, df = 1 (P = 0.62); I^2 = 0%

Test for overall effect: Z = 1.32 (P = 0.19)

Five trials reported data on the effect of eHealth on depression (1600 participants) [8,24,25,28,29]. We found no difference in effect of eHealth interventions compared with no eHealth interventions (SMD –0.13; 95% CI –0.37 to 0.11; I^2=78% [Figure S1, Multimedia Appendix 4]). However, as our main analysis showed high statistical heterogeneity, we decided to explore this in a post hoc sensitivity analysis by excluding one small trial with extreme results and a remarkably low standard deviation that also measured depression on a different scale (Patient-Reported Outcomes Measurement Information System) [28] than the other 4 trials, which measured depression on the same scale (Center for Epidemiologic Studies Depression Scale). In our sensitivity analysis (1578 participants), we also found no effect of eHealth interventions compared with no eHealth interventions; however, the heterogeneity disappeared (MD –0.73; 95% CI –2.61 to 1.16; I^2: 0% [Figure S2, Multimedia Appendix 4]). Three trials (1267 participants) reported data on PTSD [8,25,29], and we found no effect of eHealth interventions compared with no eHealth interventions (MD –0.11; 95% CI –1.04 to 0.82; I^2=0% [PTSD Checklist; Figure S3, Multimedia Appendix 4]).

Subgroup Analyses

We conducted a number of prespecified subgroup analyses on our primary outcome, overall IPV. Our subgroup analysis that compared low risk with high risk of bias trials showed similar results as our primary analysis on overall IPV ($\text{SMD}_{\text{low risk bias}}$ –0.03; 95% CI –0.15 to 0.10; I^2=0% versus $\text{SMD}_{\text{high risk bias}}$ 0.01; 95% CI –0.15 to 0.16; I^2=0%; interaction test P=.75 [Figure

S4, Multimedia Appendix 4]). Similarly, our subgroup analyses showed no effect of eHealth on reduction of IPV, if data was stratified according to type of scale (SMD [Women's Experience With Battering Scale] –0.03; 95% CI –0.15 to 0.10; I^2=0% vs SMD [Composite Abuse Scale] 0.01; 95% CI –0.15 to 0.16; I^2=0%; interaction test *P*=.75 [Figure S5, Multimedia Appendix 4]), or if data were stratified according to type of eHealth intervention ($\text{SMD}_{\text{telephone support}}$ 0.06; 95% CI –0.19 to 0.31; I^2=NA vs $\text{SMD}_{\text{online decision aid}}$ –0.02; 95% CI –0.12 to 0.09; I^2=0% vs $\text{SMD}_{\text{online education}}$ –0.24; 95% CI –0.79 to 0.30; I^2=NA; interaction test *P*=.59 [Figure S6, Multimedia Appendix 4]). One study of online education had a point estimate of relevant effect size in the favorable direction [27]; however, the confidence interval was wide and the study was judged as high risk of bias.

Discussion

Principal Findings

In this systematic review and meta-analysis, we found no evidence that eHealth interventions reduced physical, sexual, or psychological violence, depression, or PTSD compared with no eHealth intervention. We explored if the effect of eHealth interventions varied between type of intervention or the IPV scale used in our subgroup analyses but found no differences. A total of 14 trials set in the United States, Australia, and New Zealand were included in our review. Of the 8 published trials, 2 were assessed as overall in low risk of bias, and 7 trials were eligible to be included in one or more of our meta-analyses. The included studies had considerable heterogeneity in terms of type of eHealth interventions, recruitment strategies, reported outcomes, and outcome measurement tools. While in most analyses this did not result in considerable statistical heterogeneity, this limited our ability to pool the results and identify patterns across studies.

Comparison With Other Literature

Our findings are somewhat in line with other systematic reviews within this field that partly include eHealth trials. A 2014 Cochrane review that included 13 trials with 3417 participants on interventions to reduce or prevent IPV among pregnant women found that there was lack of consistency in the reported outcomes and therefore meta-analysis was not undertaken [14]. A 2015 Cochrane review that included 13 trials with 2141 participants on advocacy interventions to reduce IPV and promote psychosocial well-being of women also found considerable heterogeneity across trials with a wide range of outcomes (n=25), measurement scales, types of interventions, and time points of outcome measurements. As a result, most of the trials could not be pooled. For the trials they did manage to pool, the authors found no evidence of effect for the majority of violence outcome. None of the studies included in the review were judged to be of good quality, and the authors concluded that it was uncertain how much advocacy interventions benefit women exposed to violence [15].

Strengths and Limitations

This is the first systematic review focusing on the effect of eHealth interventions on IPV. We conducted a comprehensive literature search that involved both published and unpublished trials. However, our study has some limitations. First, we were unable to include data from unpublished trials and one published trial in our meta-analysis. Nevertheless, our effect estimate was precise for our primary analysis (ie, CI 95% –0.11 to 0.02); hence, a clinically meaningful effect of eHealth interventions on overall IPV appears to be minimal. Second, all the published trials were conducted in high-income countries, which may limit the generalizability. It is plausible that both the attitude toward IPV and the adaptation to eHealth interventions may be affected by local conditions [12]. Third, we chose to limit our meta-analysis to the outcomes overall IPV; physical, sexual, and psychological violence; depression; and PTSD as we saw these outcomes as clinically most relevant. However, our results show that these outcomes were not reported in all trials or necessarily one of the primary outcomes selected by the trial authors. Therefore, it might have been relevant to analyze other proxy measures for IPV (eg, safety behavior or self–efficacy) or other types of violence (eg, financial violence). However, including additional outcomes would increase the risk of a type I error. Fourth, outcome reporting was generally poor, and we found that 4 of the 8 published trials did not have a trial registry record; further, 1 trial did not report all prespecified outcomes. This leaves a concern for selective reporting where outcomes are selected based on the direction of findings. However, such bias generally leads to overestimation of intervention effects and therefore is unlikely to influence our conclusions. Finally, a limited number of trials were eligible to be included in this review. Hence, a future scoping review with broader eligibility criteria may complement this review and provide a more comprehensive understanding of the current state of the literature.

Implications for Practice and Research

Based on this review, we recommend conducting more high-quality trials within the field of IPV and eHealth to better ascertain the effect of eHealth interventions on IPV and IPV-related outcomes. While we found no effect of eHealth interventions despite their potential to provide a safe space for survivors, it is plausible that eHealth interventions cannot stand alone as an intervention to overcome a complex issue such as IPV. Future research may consider assessing the effect of eHealth in combination with other interventions.

The serious issue of heterogeneity in relation to types of outcomes and outcome measurements in IPV trials suggests that there is currently no consensus on which outcomes are important and how to measure them within in the field. This problem appears to go beyond eHealth and be a general problem within IPV intervention trials [14,15]. Other clinical areas have had similar issues in relation to lack of uniform outcomes, and this has led to initiatives that aim to establish core outcome sets within the fields. The Core Outcomes in Women's and Newborn Health initiative is an international initiative led by journal editors to harmonize outcome reporting in women's health research [38,39]. It is part of the Core Outcome Measures in

Effectiveness Trials initiative that strives to develop core outcome sets for clinical trials and other types of research [40]. Similarly, the Outcome Measures in Rheumatology initiative has led to the development of core outcome sets in rheumatology [41]. With inspiration from other clinical areas, we therefore recommend establishing an initiative within IPV that strives to develop core outcome sets that as a minimum should be measured and reported within IPV research.

Conclusions

This systematic review and meta-analysis found no evidence from randomized trials of a beneficial effect of eHealth interventions on overall IPV; physical, sexual, or psychological violence; or depression and PTSD. However, the types of outcomes and how they were measured were very heterogenous across trials, which limited the possibility of pooling results and identifying patterns across studies. More high-quality trials are needed, and we recommend harmonizing outcome reporting in IPV trials by establishing core outcome sets.

Acknowledgments

Special thanks to the University of Southern Denmark’s librarian, Mette Brandt Eriksen, for assisting in developing the database search strings. All authors were financed through their institutions. No external funding was received for this study. The institutions were not involved in the study and had no role in developing the study design, conduct, analysis, reporting, or decision to submit for publication.

Authors' Contributions

DSL, AB, AL, and VR conceptualized the study and wrote the protocol. DSL, AB and ANN curated the data. DSL analyzed the data, and AL and NBH verified it. DSL drafted the manuscript, and AB, ANN, AL, NBH, and VR critically revised it. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.
[PDF File (Adobe PDF File), 92 KB - jmir_v22i12e22361_app1.pdf]

Multimedia Appendix 2
Database search strings.
[PDF File (Adobe PDF File), 173 KB - jmir_v22i12e22361_app2.pdf]

Multimedia Appendix 3
Risk of bias assessment.
[PDF File (Adobe PDF File), 129 KB - jmir_v22i12e22361_app3.pdf]

Multimedia Appendix 4
Additional meta-analyses.
[PDF File (Adobe PDF File), 995 KB - jmir_v22i12e22361_app4.pdf]

References

1. Violence against women. Geneva: World Health Organization; 2019. URL: https://www.who.int/news-room/fact-sheets/detail/violence-against-women [accessed 2019-11-20]
2. García-Moreno C, Hegarty K, d'Oliveira AFL, Koziol-McLain J, Colombini M, Feder G. The health-systems response to violence against women. Lancet 2015 Apr 18;385(9977):1567-1579. [doi: 10.1016/S0140-6736(14)61837-7] [Medline: 25467583]
3. Ellsberg M, Arango DJ, Morton M, Gennari F, Kiplesund S, Contreras M, et al. Prevention of violence against women and girls: what does the evidence say? Lancet 2015 Apr 18;385(9977):1555-1566. [doi: 10.1016/S0140-6736(14)61703-7] [Medline: 25467575]
4. Global status report on violence prevention. Geneva: World Health Organization; 2014. URL: http://apps.who.int/iris/bitstream/10665/145086/1/9789241564793_eng.pdf?ua=1&ua=1 [accessed 2020-11-17]
5. Krug EG, Mercy JA, Dahlberg LL, Zwi AB. The world report on violence and health. Lancet 2002 Oct 05;360(9339):1083-1088. [doi: 10.1016/S0140-6736(02)11133-0] [Medline: 12384003]

6. Smith S, Zhang X, Basile K, Merrick M, Wang J, Kresnow M. National Intimate Partner and Sexual Violence Survey (NISVS): 2015 data brief—updated release. Atlanta: Centers for Disease Control and Prevention; 2018. URL: https://www.cdc.gov/violenceprevention/pdf/2015data-brief508.pdf [accessed 2020-11-23]
7. Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva: World Health Organization; 2013. URL: https://apps.who.int/iris/rest/bitstreams/302733/retrieve [accessed 2020-11-23]
8. Glass NE, Perrin NA, Hanson GC, Bloom TL, Messing JT, Clough AS, et al. The longitudinal impact of an internet safety decision aid for abused women. Am J Prev Med 2017 May;52(5):606-615. [doi: 10.1016/j.amepre.2016.12.014] [Medline: 28108189]
9. Campbell JC. Health consequences of intimate partner violence. Lancet 2002 Apr 13;359(9314):1331-1336. [doi: 10.1016/S0140-6736(02)08336-8] [Medline: 11965295]
10. Global plan of action to strengthen the role of the health system within a national multisectorial response to address interpersonal violence, in particular against women and girls, and against children. Geneva: World Health Organization; 2016. URL: http://apps.who.int/iris/bitstream/10665/252276/1/9789241511537-eng.pdf?ua=1 [accessed 2020-11-23]
11. eHealth at WHO. Geneva: World Health Organization URL: https://www.who.int/ehealth/about/en/ [accessed 2019-09-02]
12. Global diffusion of eHealth: making universal health coverage achievable. Report of the Third Global Survey on eHealth. Geneva: World Health Organization; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf?sequence=1&isAllowed=y [accessed 2020-11-23]
13. Westbrook L. Understanding crisis information needs in context: the case of intimate partner violence survivors. Library Q 2008 Jul;78(3):237-261. [doi: 10.1086/588443]
14. Jahanfar S, Howard LM, Medley N. Interventions for preventing or reducing domestic violence against pregnant women. Cochrane Database Syst Rev 2014 Nov 12(11):CD009414 [FREE Full text] [doi: 10.1002/14651858.CD009414.pub3] [Medline: 25390767]
15. Rivas C, Ramsay J, Sadowski L, Davidson LL, Dunne D, Eldridge S, et al. Advocacy interventions to reduce or eliminate violence and promote the physical and psychosocial well-being of women who experience intimate partner abuse. Cochrane Database Syst Rev 2015(12):CD005043. [doi: 10.1002/14651858.CD005043.pub3] [Medline: 26632986]
16. Curry MA, Durham L, Bullock L, Bloom T, Davis J. Nurse case management for pregnant women experiencing or at risk for abuse. J Obstet Gynecol Neonatal Nurs 2006;35(2):181-192. [doi: 10.1111/j.1552-6909.2006.00027.x] [Medline: 16620243]
17. Tiwari A, Fong DYT, Yuen KH, Yuk H, Pang P, Humphreys J, et al. Effect of an advocacy intervention on mental health in Chinese women survivors of intimate partner violence: a randomized controlled trial. JAMA 2010 Aug 04;304(5):536-543. [doi: 10.1001/jama.2010.1052] [Medline: 20682933]
18. Gillum TL, Sun CJ, Woods AB. Can a health clinic-based intervention increase safety in abused women? Results from a pilot study. J Womens Health (Larchmt) 2009 Aug;18(8):1259-1264 [FREE Full text] [doi: 10.1089/jwh.2008.1099] [Medline: 19627223]
19. Ehealth interventions targeted at women exposed to intimate partner violence: a systematic review of randomised trials. National Institute for Health Research. URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=130124 [accessed 2019-05-25]
20. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
21. Covidence. URL: https://www.covidence.org/ [accessed 2020-09-15]
22. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0. 2011. URL: https://training.cochrane.org/cochrane-handbook-systematic-reviews-interventions [accessed 2020-11-23]
23. Mayo-Wilson E, Li T, Fusco N, Bertizzolo L, Canner JK, Cowley T, et al. Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy. J Clin Epidemiol 2017 Nov;91:95-110 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.07.014] [Medline: 28842290]
24. Hegarty K, Tarzia L, Valpied J, Murray E, Humphreys C, Taft A, et al. An online healthy relationship tool and safety decision aid for women experiencing intimate partner violence (I-DECIDE): a randomised controlled trial. Lancet Public Health 2019 Jun;4(6):e301-e310 [FREE Full text] [doi: 10.1016/S2468-2667(19)30079-9] [Medline: 31155223]
25. Koziol-McLain J, Vandal AC, Wilson D, Nada-Raja S, Dobbs T, McLean C, et al. Efficacy of a web-based safety decision aid for women experiencing intimate partner violence: randomized controlled trial. J Med Internet Res 2018 Jan 10;19(12):e426 [FREE Full text] [doi: 10.2196/jmir.8617] [Medline: 29321125]
26. Eden KB, Perrin NA, Hanson GC, Messing JT, Bloom TL, Campbell JC, et al. Use of online safety decision aid by abused women: effect on decisional conflict in a randomized controlled trial. Am J Prev Med 2015 Apr;48(4):372-383 [FREE Full text] [doi: 10.1016/j.amepre.2014.09.027] [Medline: 25547929]
27. Zlotnick C, Tzilos Wernette G, Raker CA. A randomized controlled trial of a computer-based brief intervention for victimized perinatal women seeking mental health treatment. Arch Womens Ment Health 2019 Jun;22(3):315-325 [FREE Full text] [doi: 10.1007/s00737-018-0895-1] [Medline: 30088145]

28. Constantino RE, Braxter B, Ren D, Burroughs JD, Doswell WM, Wu L, et al. Comparing online with face-to-face HELPP intervention in women experiencing intimate partner violence. Issues Ment Health Nurs 2015 Jun;36(6):430-438. [doi: 10.3109/01612840.2014.991049] [Medline: 26241569]
29. Stevens J, Scribano PV, Marshall J, Nadkarni R, Hayes J, Kelleher KJ. A trial of telephone support services to prevent further intimate partner violence. Violence Against Women 2015 Dec;21(12):1528-1547. [doi: 10.1177/1077801215596849] [Medline: 26223696]
30. McFarlane J, Malecha A, Gist J, Watson K, Batten E, Hall I, et al. An intervention to increase safety behaviors of abused women: results of a randomized clinical trial. Nurs Res 2002;51(6):347-354. [doi: 10.1097/00006199-200211000-00002] [Medline: 12464754]
31. Braithwaite SR, Fincham FD. Computer-based prevention of intimate partner violence in marriage. Behav Res Ther 2014 Mar;54:12-21. [doi: 10.1016/j.brat.2013.12.006] [Medline: 24463577]
32. Ford-Gilboe M, Varcoe C, Scott-Storey K, Wuest J, Case J, Currie LM, et al. A tailored online safety and health intervention for women experiencing intimate partner violence: the iCAN Plan 4 Safety randomized controlled trial protocol. BMC Public Health 2017 Mar 21;17(1):273 [FREE Full text] [doi: 10.1186/s12889-017-4143-9] [Medline: 28327116]
33. Sabri B, Njie-Carr VPS, Messing JT, Glass N, Brockie T, Hanson G, et al. The weWomen and ourCircle randomized controlled trial protocol: a web-based intervention for immigrant, refugee and indigenous women with intimate partner violence experiences. Contemp Clin Trials 2019 Jan;76:79-84 [FREE Full text] [doi: 10.1016/j.cct.2018.11.013] [Medline: 30517888]
34. Henriksen L, Flaathen EM, Angelshaug J, Garnweidner-Holme L, Småstuen MC, Noll J, et al. The Safe Pregnancy study—promoting safety behaviours in antenatal care among Norwegian, Pakistani and Somali pregnant women: a study protocol for a randomized controlled trial. BMC Public Health 2019 Jun 10;19(1):724 [FREE Full text] [doi: 10.1186/s12889-019-6922-y] [Medline: 31182062]
35. Clark CJ, Spencer RA, Shrestha B, Ferguson G, Oakes JM, Gupta J. Evaluating a multicomponent social behaviour change communication strategy to reduce intimate partner violence among married couples: study protocol for a cluster randomized trial in Nepal. BMC Public Health 2017 Jan 13;17(1):75 [FREE Full text] [doi: 10.1186/s12889-016-3909-9] [Medline: 28086857]
36. SAFE: a self-support eHealth intervention to support women exposed to intimate partner violence (NTR7313). Amsterdam: Netherlands Trial Register URL: https://www.trialregister.nl/trial/7108 [accessed 2020-09-15]
37. MyPlan4SexualSafety: synergising eHealth and behavioral science to promote safety for women (PACTR04003321122). Cape Town: Pan-African Clinical Trials Registry URL: https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=3321 [accessed 2020-11-23]
38. Duffy J, Rolph R, Gale C, Hirsch M, Khan KS, Ziebland S, International Collaboration to Harmonise Outcomes in Pre-eclampsia (iHOPE). Core outcome sets in women's and newborn health: a systematic review. BJOG 2017 Sep;124(10):1481-1489 [FREE Full text] [doi: 10.1111/1471-0528.14694] [Medline: 28421657]
39. Cooper N, Duffy J. CROWN: Core outcomes in women's and newborn health. URL: http://www.crown-initiative.org/ [accessed 2020-06-09]
40. Wiliamson P, Gargon E, Bagley H, Blazeby J, Clarke M, Tunis S. COMET initiative: core outcome measures in effectiveness trials. URL: http://www.comet-initiative.org/ [accessed 2020-06-09]
41. The OMERACT handbook. URL: https://omeracthandbook.org/ [accessed 2020-06-09]

Abbreviations

IPV: intimate partner violence
MD: mean difference
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register for Systematic Reviews
PTSD: posttraumatic stress disorder
SMD: standardized mean difference

Edited by G Eysenbach; submitted 09.07.20; peer-reviewed by L Genaro, K Mokgalaboni, L Schwab-Reese, J Feliciano; comments to author 04.09.20; revised version received 15.09.20; accepted 15.09.20; published 11.12.20.

Please cite as:
Linde DS, Bakiewicz A, Normann AK, Hansen NB, Lundh A, Rasch V
Intimate Partner Violence and Electronic Health Interventions: Systematic Review and Meta-Analysis of Randomized Trials
J Med Internet Res 2020;22(12):e22361
URL: http://www.jmir.org/2020/12/e22361/
doi:10.2196/22361
PMID:33306030

Review

Technical Aspects of Developing Chatbots for Medical Applications: Scoping Review

Zeineb Safi[1], MSc; Alaa Abd-Alrazaq[1], PhD; Mohamed Khalifa[2], PhD; Mowafa Househ[1], PhD

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

[2]Centre for Health Informatics, Australian Institute of Health Innovation, Faculty of Medicine, Health, and Human Sciences, Macquarie University, Sydney, Australia

Corresponding Author:
Mowafa Househ, PhD
Division of Information and Computing Technology, College of Science and Engineering
Hamad Bin Khalifa University
Qatar Foundation
P.O. Box 34110
Doha Al Luqta St, Ar-Rayyan
Doha
Qatar
Phone: 974 55708549
Email: mhouseh@hbku.edu.qa

Abstract

Background: Chatbots are applications that can conduct natural language conversations with users. In the medical field, chatbots have been developed and used to serve different purposes. They provide patients with timely information that can be critical in some scenarios, such as access to mental health resources. Since the development of the first chatbot, ELIZA, in the late 1960s, much effort has followed to produce chatbots for various health purposes developed in different ways.

Objective: This study aimed to explore the technical aspects and development methodologies associated with chatbots used in the medical field to explain the best methods of development and support chatbot development researchers on their future work.

Methods: We searched for relevant articles in 8 literature databases (IEEE, ACM, Springer, ScienceDirect, Embase, MEDLINE, PsycINFO, and Google Scholar). We also performed forward and backward reference checking of the selected articles. Study selection was performed by one reviewer, and 50% of the selected studies were randomly checked by a second reviewer. A narrative approach was used for result synthesis. Chatbots were classified based on the different technical aspects of their development. The main chatbot components were identified in addition to the different techniques for implementing each module.

Results: The original search returned 2481 publications, of which we identified 45 studies that matched our inclusion and exclusion criteria. The most common language of communication between users and chatbots was English (n=23). We identified 4 main modules: text understanding module, dialog management module, database layer, and text generation module. The most common technique for developing text understanding and dialogue management is the pattern matching method (n=18 and n=25, respectively). The most common text generation is fixed output (n=36). Very few studies relied on generating original output. Most studies kept a medical knowledge base to be used by the chatbot for different purposes throughout the conversations. A few studies kept conversation scripts and collected user data and previous conversations.

Conclusions: Many chatbots have been developed for medical use, at an increasing rate. There is a recent, apparent shift in adopting machine learning–based approaches for developing chatbot systems. Further research can be conducted to link clinical outcomes to different chatbot development techniques and technical characteristics.

(J Med Internet Res 2020;22(12):e19127) doi:10.2196/19127

KEYWORDS

chatbots; conversational agents; medical applications; scoping review; technical aspects

Introduction

Chatbots are systems that are capable of conversing with users in natural language in a way that simulates the interaction with a real human. The development of chatbots has captured the attention of researchers for a long time. Eliza [1] was one of the earliest attempts at developing a conversational system. Since then, chatbot development has evolved to be an integral part of many application domains. The most prominent example is the use of chatbots as personal assistants such as Apple's Siri and Google Assistant. Chatbots have also been developed and are being used in different application areas, such as marketing and to provide different types of services [2,3].

Since the early years of their development, people adopted different techniques in designing and developing chatbots. In recent years, with the increasing adoption of machine learning and artificial intelligence methods in different application domains, state-of-the-art methods in dialogue generation and dialogue management are increasingly using machine learning and deep learning methods [4-6].

The medical field is an application area where chatbots are increasingly being adopted as a tool to facilitate access to information from the patient side and reduce the load on the clinician side. Many commercial chatbot applications that are available as web or mobile applications have been developed for interacting with patients. Some examples of health care chatbots are OneRemission1, which was developed to help cancer survivors; Babylon Health, which is a symptom checker; and Wysa [7], which is a mental health chatbot that interacts with user to help with signs of anxiety and depression.

It is important to know the current state of different methods and techniques that are being employed in developing chatbots in the medical domain for many reasons. Conducting such a survey will help researchers in the future identify the different methods that have been used and to build on the existing approaches to develop more intelligent chatbots that provide a more natural experience to the user. It is also important to see where the current state of chatbot development stands with respect to developing chatbots for other applications. Therefore, in this work, we conducted a scoping review of the available literature on chatbot development in the medical field and constructed and identified the main components involved in chatbot development, as well as a description of techniques used in developing each of the identified components. The main objective of this study was to explore technical aspects and development methodologies associated with chatbots used in the medical field to explain best methods of development and support chatbot development researchers in their future work.

Methods

This study follows a scoping review methodology. Specifically, it follows the PRISMA extension of scoping reviews [8]. In this section, we explain the details of the adopted methodology to conduct the review. The PRISMA extension for scoping reviews is presented in Multimedia Appendix 1.

Search Strategy

Search Sources

We searched 8 databases (IEEE, ACM, Springer, ScienceDirect, Embase, MEDLINE, PsycINFO, and Google Scholar) to collect studies relevant to the topic. For Google Scholar, we only used the first 100 results from each search string, as Google Scholar returns the most relevant results belonging to each search query first. The search was conducted between September 9, 2019 and September 13, 2019. For the forward reference list checking, we used the cited-by functionality of Google scholar. We also checked the reference list of the included studies to review the backward reference list.

Search Terms

We used 2 different sets of search terms to search the databases. The search term set was decided based on the type of studies indexed by the database. For databases that mainly indexed studies in the medical field (Embase, MEDLINE, PsycINFO), we relied on keywords that are strictly related to the intervention (eg, chatbot, chatterbot, conversational agent, conversational bot). For databases with no specific application domain, we resorted to using keywords that are related to the medical domain (eg, health, medical, illness, disease, disability) in addition to the intervention-related words. The search strategy used for searching the databases is presented in Multimedia Appendix 2.

Study Eligibility Criteria

The purpose of this work was to review the technical aspects of developing text-based chatbots in the medical field. Therefore, for a study to be considered, it had to satisfy the following criteria: describe a chatbot application, the chatbot must be developed for a medical application (eg, management, diagnosis, counseling), the input or output modality of the chatbot must be text, and the technical details of how the input is processed and output is produced must be mentioned. Studies that used a Wizard of Oz experiment design were excluded. In addition, some restrictions on the language of the study and publication type were enforced. Only studies that were published in English were included, and only peer-reviewed articles, conference papers, thesis, dissertations, and industrial and academic reports were considered.

Study Selection

The study selection was conducted in 2 stages. Title and abstract screening was followed by a full-text screening stage. Both stages were conducted by 2 reviewers. The first reviewer, ZS, performed the screening of the full set of articles. Due to time constraints, the second reviewer, AA, reviewed a randomly selected set of 50% of the articles. Disagreements between the reviewers were resolved by a third reviewer, MH. To evaluate the interrater agreement, we used Cohen kappa [9]. The reviewers had substantial agreement in both stages, with a kappa measure of 0.74 and 0.67 in the first and second stages, respectively.

Data Extraction

The data extraction was conducted by ZS following a preset form. The data extracted pertained to the metadata of the included studies as well as the different technical modules of interest in the study, such as the text understanding module, text generation module, and method of linking these modules. The data extraction form is shown in Multimedia Appendix 3.

Study Quality Assessment

As this is a scoping, not a systematic, review, no study quality assessment was conducted for the purposes of this work.

Data Synthesis

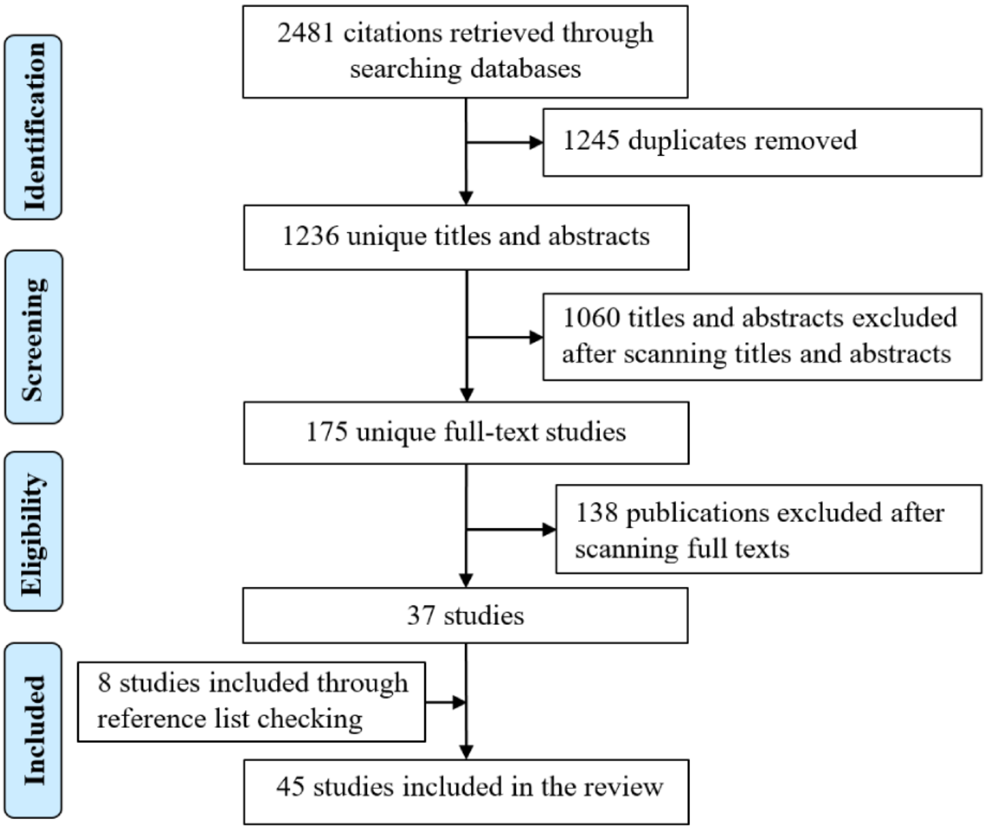
We used a narrative approach to synthesize the different reported results. We included a description of the included studies and a description of the different techniques used for the development of the chatbots.

Results

Search Results

Figure 1 summarizes the process that was followed to select the studies. Of the 2481 total studies returned after searching the databases, 1245 were duplicated. After removing the duplicates, 1236 studies remained and were screened based on title and abstract. After the title and abstract–based screening, 1060 studies were removed for the following reasons: not describing a chatbot (n=840), not containing technical details of the chatbot implementation (n=4), not belonging to a medical application (n=172), not containing text understanding or text generation (n=5), not written in the English language (n=8), and non-peer–reviewed publications (n=31). After the full-text screening phase, 138 additional studies were removed for the following reasons: not describing a chatbot (n=35), not containing technical details of the chatbot implementation (n=56), not belonging to a medical application (n=3), not containing text understanding or text generation (n=27), not written in the English language (n=1), and non-peer–reviewed publications (n=16). After performing forward and backward reference checking, 8 additional studies were included. The total number of included studies was 45.

Figure 1. Study selection process.



Description of Included Studies

The included studies were published between the years 2009 and 2019, with 80% (36/45) of the studies published in the last 5 years, in the years 2015-2019. Conference papers were the most common publication type (23/45, 51%), followed by journal papers (19/45, 42%) in addition to 2 magazine articles and 1 thesis. The most common country for publications was the United States, with 8 studies, followed by Australia, India, and Switzerland, with 4 studies each. Multimedia Appendix 4 shows the characteristics of each included study.

Description of Chatbots

The total number of chatbots assessed in this study was 45. The chatbots were developed for different diseases and to fulfill different purposes. Table 1 shows the different categories and number of chatbots developed in each category. The most common category was “General Health,” which includes chatbots designed for health educational or counseling purposes for general health conditions. These chatbots can also provide information on general medical services such as disease diagnosis based on patient-given symptoms. Some provide patients with discharge information before leaving the hospital. The second most common type of chatbot is those developed for mental health purposes, followed by those developed for specific diseases, such as diabetes, cancer, autism, heart disease, and asthma.

An important factor for the technical aspects of developing chatbots is the language that the chatbot uses for communication. The majority of chatbots communicate in English (23/45, 51.1%), 4 in German, 3 in Chinese, 2 in Arabic, 2 in Korean, 1 in Thai, 1 in Spanish, and 1 in Russian; 8 studies did not mention the language. The chatbots operated as either standalone applications (n=17) or web applications (n=20), while the remaining studies did not mention the application type (n=8).

Table 1. Target diseases for chatbot development in the included studies (n=45).

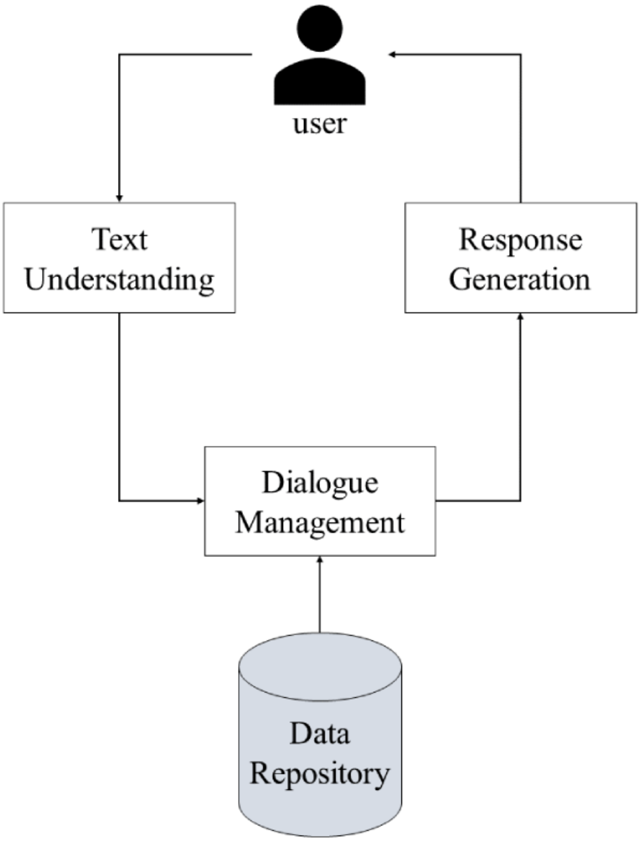
Disease/condition	Count	Percentage (%)	Studies
General health	21	47	[10-31]
Mental health	15	33	[32-45]
Diabetes	2	4	[46,47]
Cancer	2	4	[48,49]
Autism	2	4	[50,51]
Heart disease	1	2	[52]
Asthma	1	2	[53]
HIV	1	2	[54]

Chatbot Implementation

Overview

The chatbots in the included studies consisted of 4 main modules: text understanding module, dialog manager, text generation module, and database layer that holds the various types of information needed for chatbot training and function. Figure 2 shows a high-level architecture of chatbot development and the relationship between the different modules that constitute it. The following subsections highlight the findings in terms of how each of the previously mentioned components is implemented.

Figure 2. High-level chatbot architecture.



Text Understanding

The text understanding module is the module with which the user directly interacts. The function of the text understanding module is to extract meaning from the user input before a specific answer is generated. In the included studies, 7 methods were used for text understanding. Pattern matching methods were the most common, used by 18 of the 45 chatbots (40%). Many text understanding approaches can fall under pattern matching. The different pattern matching methods reported in the included studies are as follows: keyword matching or string matching, where specific keywords or strings of the user input are matched to scripts stored in the database, which was used in 11 studies [10-15,32,33,46,48,50], or Extensible Markup Language (XML) variants, such as Artificial Intelligence Markup Language (AIML), which were used in 7 studies [16,17,34-37,47]. The second most popular approach was the use of machine learning methods for text understanding, which was used in 6 studies. Most of these studies used supervised machine learning algorithms, including decision trees and random forests. These include groups of decision trees, where each tree gives a classification to the text, in the form of voting, while the forest chooses the classification having the most votes over all the trees in the forest. Machine learning training materials were sometimes based on real health data extracted and anonymized from hospital information systems and electronic health records. In other instances, chatbots were trained on billions of words extracted from Google News datasets [18,19,38,39,49,51]. Another 6 studies relied on web services such as Dialogflow by Google. Web services are usually provided through a computer server that responds to requests through ports over a network, such as the internet, to solve specific domain problems over the web. They can also use a mixture of machine learning and rule-based methods to produce a customizable chatbot implementation. Three studies used natural language processing–based approaches, such as named entity recognition, to extract meaning from user input. To develop the chatbot, 2 of the studies used a rule-based approach, where the chatbot is operated by either a set of IF-THEN rules or a state-based system. A few of the studies relied on a fixed input method, where the user selects the input from a list of possible inputs. Fixed input chatbots do not technically need to go through a text understanding module as the system does not need to interpret the user input. The input needs only to be directed to the dialogue manager. Some studies proposed a few hybrid text understanding approaches. Hybrid approaches use a combination of 2 or more of the previously mentioned methods for text understanding. Table 2 summarizes the different text understanding methods proposed and the studies that fall under each method.

Table 2. Text understanding methods used in the studies (n=45).

Text understanding method	Studies
Pattern matching	[10-17,32-37,46-48,50]
Machine learning	[18,19,38,39,49,51]
Web services	[20-22,40,53,54]
Fixed input	[23-25,41,42]
Natural language processing	[26-28]
Hybrid	[43,44,52]
Rule-based	[29,30]
Not mentioned	[31,45]

Dialogue Management

In the reported studies, the input to the dialog management module is the processed user free-text input provided by the text understanding module. The dialog management module controls the different aspects of the conversation and links each user input to an appropriate output. In the included studies, 2 main types of dialog management techniques were reported: static dialog management and dynamic dialog management. In the case of static dialog management, user input is matched directly to the appropriate output using a pattern matching algorithm (25/45) or using a set of rules (7/45). In the case of dynamic dialogue management, the context of the conversation changes based on specific user input characteristics. The context switching can be done by training machine learning algorithms to identify the context from the user input (5/45) or using a web service for intent identification (4/45). Table 3 summarizes the different dialogue management methods and the studies that correspond to each method.

Table 3. Dialogue management methods in the included studies (n=45).

Dialogue management method/platform	Studies
Pattern matching algorithms	[10,12,14-19,22,25,28,32-38,42,45-50]
Rule-based methods	[11,26,27,29-31,52]
Machine learning	[39,40,43,44,51]
Web-based	[20,21,53,54]
Not mentioned	[13,23,24,41]

Data Management

Most of the chatbots in the included studies contained one or more of the 3 data repository types identified. Most of the included studies kept a medical knowledge repository (28/45, 62%). The medical knowledge repository contains medical information related to the application domain of the developed chatbot. The medical knowledge source can be local, such as obtaining it from medical personnel, or it can be collected from online sources such as Wikipedia or other websites. The medical knowledge can be presented to the user in the context of educational chatbots (17/45) [12-14,19,23-25,27,32,34,41,46,48-51,53], or it can be used to train machine learning algorithms (5/45) [39,40,43,44,51]. Many of the developed approaches store users' data and use the data to customize the chatbot response and improve its functionality (11/45). Conversation scripts are the third-data repository type, and they are usually kept by chatbots that use pattern matching as a text understanding or text generation modality (5/45). A few studies did not mention the type of data stored (9/45). Table 4 summarizes the different database types and studies that reported keeping each database type.

Table 4. Database types in the included studies (n=45).

Database type	Studies
Medical knowledge database	[10,11,13,15-17,20,22-24,26-28,31-33,36-42,47,48,50-53]
User information database	[11,13,18,21,25,39,40,45,46,50,53]
Conversation scripts	[11,12,14,15,19]
Not mentioned	[21,29,30,34,35,43,44,49,54]

Text Generation

The text generation module provides output to the user. Text generation in the included studies was done using one of 2 methods: fixed output or generated output. The fixed output methods search the database for the most appropriate output to a user input and present it to the user. The generated output method relies on machine learning to generate original natural language output that is produced by the machine learning algorithm. Chatbots in most of the included studies provide fixed output that is extracted from the database with the exception of those in [25,39,51], which provide output that is generated using machine learning and deep learning methods, and [43], which proposes a hybrid approach that can provide fixed and generated outputs. A few studies did not report the output generation method, or it was not applicable as the output modality was not text [23,24,41,42,45].

Discussion

Principal Findings

A general architecture was identified and reported to summarize the technical aspects of chatbot development. The main components of chatbots, as well as the way these components are linked, are reported. Chatbots typically consist of 4 main components: text understanding module, dialogue management module, data management layer, and text generation module.

The most common design method employed in developing chatbots is pattern matching for text understanding and response generation. Machine learning and generative methods are among the least commonly used methods for the development of chatbots in the medical domain. This can be attributed to 2 main reasons. The first reason for relying on pattern matching approaches more than those based on machine learning is that pattern matching methods are more reliable in practice because they produce exact responses to well-defined queries, resulting in fewer mistakes. Machine learning–based methods usually produce different types of errors, which cannot be tolerated in medical applications. The second reason for this trend is the rapid development in the state of the machine learning field over the past few years and the increase in the robustness of its methods, especially with the emergence of deep learning. While older methods relied on rule-based chatbots and pattern matching algorithms, all the proposed methods that rely on machine learning for text understanding and response generation were proposed between the years 2017 and 2019. Another reason for the possible lack of using machine learning methods could be the fact that machine learning–based approaches need to be trained using large amounts of domain-specific data, which might be scarce and difficult to access in the medical field. Overall, machine learning approaches and algorithms were better suited for developing chatbots used for specific medical conditions, such as mental health and autism, while the rule-based approaches were better suited for developing chatbots used for general medical purposes. On the other hand, pattern matching methods and algorithms were more broadly used in developing chatbots used for both special and general medical conditions.

In terms of data management, the developed chatbots kept track of 3 different types of databases: medical knowledgebase including a library of medical facts, user information database including details about users' demographics and their preferences, and dialogue script database including all possible entries of conversational text responding to users. The type of database kept depends on the chatbot type and target functionality. Educational chatbots usually keep a medical knowledgebase. Chatbots that use context switching based on user emotions usually keep a user information database.

Most of the developed chatbots used English as the language of communication with the users, while other languages such as German, Chinese, and Arabic were less common. This is consistent with the fact that most of the publications originated in the United States, followed by Australia, where the first language is English.

Strengths and Limitations

Strengths

This review focused generally on chatbots in the medical field, without specifying the field of application, which makes it more comprehensive than previously conducted reviews [55]. Other reviews that included different medical applications [56] presented a general taxonomy of conversational agents, while we presented a more granular description of the development techniques of each component. Searching more libraries from different application domains allowed us to include more chatbots in our study than that of Montenegro et al [56]. This is why we excluded Wizard of Oz studies, while they were used in the previous reviews. Two similar systematic review studies were conducted earlier in 2018 and 2019 [57,58]. The first study reviewed the applications and evaluation measures of chatbots and conversational agents, while the second paper provided a critical review of the tasks involved in natural language understanding and machine learning of chatbot systems used in the medical domain. Neither of the 2 papers discussed the technical aspects and development methodologies of the chatbots used in the medical domain.

Limitations

This review only focused on text-based chatbot applications, where either the input or output modality is written. This excludes studies where the input or output modalities are spoken or visual, as well as robotics and telephone-based methods. This choice was made because we wanted to focus on text processing techniques rather than image or voice processing, as speech-to-text technologies can also introduce errors and another layer of complexity to chatbot development.

We enforced some constraints on the type of publications that were included in the current review. These constraints might have led to missing a portion of developed chatbots that have been published in other research venues, such as workshops, book chapters, and conference abstracts. Furthermore, limiting the search to papers published in English could also have led to missing some chatbots that were developed for communication in other languages and published in their own language. For example, we did not include papers published in Chinese or Arabic that discuss chatbots communicating in these languages.

This review focused on the development process of chatbots without considering the impact of these methods on patients. For this reason, some of the implementations in some of the included studies might be feasible from a technical point of view, but this does not necessarily mean they are effective from a medical point of view.

Practical and Research Implications

Practical Implications

This paper reports the technical aspects of developing chatbots in the medical field. This review can be used to identify the most common development approaches by specialists to help them narrow down their options and make a decision on which development approach is the most appropriate for their applications.

The reported results show that most of the developed chatbots communicate with users in the English language. While a few attempts to design and develop chatbots in other languages exist, more work needs to be done in this regard, especially for languages that are spoken by a large portion of the world's population, such as Chinese and Arabic.

Even though dynamic dialogue management provides a more natural user experience, most developed systems rely on static dialogue management methods. Changing the dialogue context based on user emotions or by detecting topic changes in user input are important aspects to be considered in chatbot development.

The use of machine learning and artificial intelligence methods in the development of conversational agents in different application areas has recently increased. The rate of adoption of machine learning–based methods in developing chatbots is still relatively low, even though it has been increasing in recent years. Supervised machine learning algorithms seem to better suit the development of chatbots for special medical conditions and diseases, while rule-based methods are being used more for developing chatbots used for general medical purposes. Machine learning methods allow the development of more intelligent agents that can provide a more realistic user experience, by providing a better text understanding experience, including more dynamic and flexible dialogue management, and generating a wider range of responses.

Research Implications

As the purpose of this review was to survey the technical aspects of chatbot development, the clinical results of performing clinical trials were not considered. Further reviews linking the different development techniques used to the clinical outcomes of the chatbot developed are possible and recommended.

More openness and a wider adoption of state-of-the-art methods in dialogue management [4], text understanding [5], and text generation [6] methods in the literature can really benefit the development of conversational agents in the medical field.

It is worth noting that the technical aspects of developing chatbots were not always clearly mentioned in the studies. The devised architecture is a general one that does not necessarily apply to every developed chatbot. One or more component might be omitted, and the chatbot might still function properly.

Conclusion

In the scope of this review, we analyzed the technical aspects of developing 45 text-based chatbots for the purpose of performing different medical interventions. The most common language used for chatbot communication is English. Chatbots typically contain 4 main components: text understanding module, dialogue manager, database layer, and text generation module. The most common technique for developing chatbots is using a string matching algorithm and a set of scripts that contain sample inputs and outputs. Judging from the publication years of the different studies, we can conclude that chatbots are becoming increasingly popular for medical application, especially when it comes to mental health. The adoption of machine learning and artificial intelligence–based techniques has recently increased. Some development approaches are better suited than others for developing chatbots for specific medical conditions rather than general medical conditions. Future studies can be conducted to link the development techniques of chatbots to their clinical outcomes. It is important to conduct more in-depth systematic reviews on the effectiveness of chatbots in supporting and enhancing positive clinical outcomes. We need to understand and correlate different technical criteria and development methodologies to different levels of chatbots acceptance, utilization, and clinical effectiveness. Discussing the pros and cons of each chatbot system has also been left to future supplementary studies, to compare advantages and disadvantages of each chatbot system and link these to their postimplementation clinical outcomes.

Acknowledgments

This study was a part of a project funded by the Qatar National Research Fund (NPRP12S-0303-190204). The project title is “A Personalized and Intelligent Digital Mental Health Platform for Qatar and the Arab World.”

Authors' Contributions

ZS developed the study protocol and conducted the search with the assistance of AA, who was the second reviewer of the studies selected. MK supported the study, revised the manuscript, improved the writing, and responded to the reviewers. The work was conducted with the guidance and supervision of MH. Study selection was done by ZS and AA. Data extraction and result synthesis were done by ZS. The manuscript was reviewed by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.
[PDF File (Adobe PDF File), 133 KB - jmir_v22i12e19127_app1.pdf]

Multimedia Appendix 2
Search strategy.
[DOCX File , 19 KB - jmir_v22i12e19127_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOC File , 14 KB - jmir_v22i12e19127_app3.doc]

Multimedia Appendix 4
Chatbot description.
[XLS File (Microsoft Excel File), 13 KB - jmir_v22i12e19127_app4.xls]

References

1. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 1966 Jan;9(1):36-45. [doi: 10.1145/365153.365168]
2. Van den Broeck E, Zarouali B, Poels K. Chatbot advertising effectiveness: When does the message get through? Computers in Human Behavior 2019 Sep;98:150-157. [doi: 10.1016/j.chb.2019.04.009]
3. Gnewuch U, Morana S, Maedche A. Towards Designing Cooperative and Social Conversational Agents for Customer Service. 2017 Presented at: International Conference on Information Systems (ICIS) 2017; December 10-13, 2017; Seoul, South Korea.
4. Serban IV, Klinger T, Tesauro G, Talamadupula K, Zhou B, Bengio Y, et al. Multiresolution Recurrent Neural Networks: An Application to Dialogue Response Generation. 2016. URL: http://arxiv.org/abs/1606.00776 [accessed 2020-02-12]
5. Henderson M, Budzianowski P, Casanueva I, Coope S, Gerz D, Kumar G, et al. A Repository of Conversational Datasets. 2019. URL: http://arxiv.org/abs/1904.06472 [accessed 2020-02-12]
6. Guo J, Lu S, Cai H, Zhang W, Yu Y, Wang J. Long Text Generation via Adversarial Training with Leaked Information. 2018 Presented at: The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18); February 2–7, 2018; New Orleans, LA.
7. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]

8. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
9. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
10. Comendador BEV, Francisco BMB, Medenilla JS, Nacion SMT, Serac TBE. Pharmabot: A Pediatric Generic Medicine Consultant Chatbot. JOACE 2015;3(2):137-140. [doi: 10.12720/joace.3.2.137-140]
11. Abashev A, Grigoryev R, Grigorian K, Boyko V. Programming Tools for Messenger-Based Chatbot System Organization: Implication for Outpatient and Translational Medicines. BioNanoSci 2016 Nov 22;7(2):403-407. [doi: 10.1007/s12668-016-0376-9]
12. Su BH, Fu PW, Wang J, Kuan T, Chiu PJ. Customized dialogue system based on biographical knowledge retrieval for elderly. 2015 Presented at: International Conference on Orange Technologies (ICOT); December 19-22, 2015; Hong Kong. [doi: 10.1109/icot.2015.7498484]
13. Tschanz M, Dorner T, Holm J, Denecke K. Using eMMA to Manage Medication. Computer 2018 Aug;51(8):18-25. [doi: 10.1109/MC.2018.3191254] [Medline: 26366060]
14. Wong W, Thangarajah J, Padgham L. Health Conversational System Based on Contextual Matching of Community-driven Question-answer Pairs. 2011 Presented at: 20th ACM International Conference on Information and Knowledge Management; October 2011; New York, NY. [doi: 10.1145/2063576.2064024]
15. Ni L, Lu C, Liu N, Liu J. MANDY: Towards a Smart Primary Care Chatbot Application. In: Chen J, Theeramunkong T, Supnithi T, Tang X, editors. Knowledge and Systems Sciences. KSS 2017. Communications in Computer and Information Science. Singapore: Springer Publishing Company; 2017:38-52.
16. Rarhi K, Mishra A, Mandal K. Automated Medical Chatbot. EasyChair. 2018 Apr 28. URL: https://easychair.org/publications/preprint/DSwT [accessed 2020-02-12]
17. Sakthivel V, Srihari K, Karthik S, Anisha CD. Intelligent Counselling Bot Using Ranking Algorithm in AI. 2018 Presented at: International Conference on Soft-computing and Network Security (ICSNS); February 14-16, 2018; Coimbatore, India. [doi: 10.1109/icsns.2018.8573642]
18. Mäurer D. Thesis: Rational Decision Support with a Natural Language Dialogue System. Technische Universität Darmstadt. 2015 Oct. URL: https://d-nb.info/1116604477/34 [accessed 2020-12-12]
19. Su MH, Wu CH, Huang KY, Hong QB, Wang HM. A chatbot using LSTM-based multi-layer embedding for elderly care. 2018 Presented at: International Conference on Orange Technologies (ICOT); December 8-10, 2017; Singapore. [doi: 10.1109/icot.2017.8336091]
20. Rosruen N, Samanchuen T. Chatbot Utilization for Medical Consultant System. 2019 Presented at: 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON); December 12-14, 2018; Bangkok, Thailand. [doi: 10.1109/times-icon.2018.8621678]
21. Fadhil A, Gabrielli S. Addressing Challenges in Promoting Healthy Lifestyles: The Al-chatbot Approach. 2017 Presented at: 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2017; Barcelona, Spain. [doi: 10.1145/3154862.3154914]
22. Fadhil A. OlloBot - Towards A Text-Based Arabic Health Conversational Agent: Evaluation and Results. 2019 Presented at: International Conference on Recent Advances in Natural Language Processing (RANLP); September 2019; Varna, Bulgaria. [doi: 10.26615/978-954-452-056-4_034]
23. Bickmore TW, Utami D, Matsuyama R, Paasche-Orlow MK. Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment. J Med Internet Res 2016 Jan 04;18(1):e1 [FREE Full text] [doi: 10.2196/jmir.5239] [Medline: 26728964]
24. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013 Aug;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]
25. Turunen M, Hakulinen J, Ståhl O, Gambäck B, Hansen P, Rodríguez Gancedo MC, et al. Multimodal and mobile conversational Health and Fitness Companions. Computer Speech & Language 2011 Apr;25(2):192-209. [doi: 10.1016/j.csl.2010.04.004]
26. Ghosh S, Bhatia S, Bhatia A. Quro: Facilitating User Symptom Check Using a Personalised Chatbot-Oriented Dialogue System. Stud Health Technol Inform 2018;252:51-56. [Medline: 30040682]
27. Wang H, Zhang Q, Ip M, Fai Lau JT. Social Media–based Conversational Agents for Health Management and Interventions. Computer 2018 Aug;51(8):26-33. [doi: 10.1109/mc.2018.3191249]
28. Chatzimina M, Koumakis L, Marias K, Tsiknakis M. Employing Conversational Agents in Palliative Care: A Feasibility Study and Preliminary Assessment. 2019 Presented at: IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE); October 28-30, 2019; Athens, Greece. [doi: 10.1109/bibe.2019.00095]
29. Huang C, Yang M, Huang C, Chen Y, Wu M, Chen K. A Chatbot-supported Smart Wireless Interactive Healthcare System for Weight Control and Health Promotion. 2019 Presented at: IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); December 16-19, 2018; Bangkok, Thailand. [doi: 10.1109/ieem.2018.8607399]

30. Kasinathan V, Xuan FS, Wahab MHA, Mustapha A. Intelligent Healthcare Chatterbot (HECIA): Case study of medical center in Malaysia. 2017 Presented at: IEEE Conference on Open Systems (ICOS); November 13-14, 2017; Miri, Malaysia. [doi: 10.1109/icos.2017.8280270]
31. Morbini F, Forbell E, DeVault D, Sagae K, Traum DR, Rizzo AA. A mixed-initiative conversational dialogue system for healthcare. 2012 Presented at: 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue; July 5-6, 2012; Seoul, South Korea URL: http://0-dl.acm.org.mylibrary.qu.edu.qa/citation.cfm?id=2392800.2392825
32. Hudlicka E. Virtual training and coaching of health behavior: example from mindfulness meditation training. Patient Educ Couns 2013 Aug;92(2):160-166 [FREE Full text] [doi: 10.1016/j.pec.2013.05.007] [Medline: 23809167]
33. Park S, Choi J, Lee S, Oh C, Kim C, La S, et al. Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study. J Med Internet Res 2019 Apr 16;21(4):e12231 [FREE Full text] [doi: 10.2196/12231] [Medline: 30990463]
34. Denecke K, May R, Deng Y. Towards Emotion-Sensitive Conversational User Interfaces in Healthcare Applications. Stud Health Technol Inform 2019 Aug 21;264:1164-1168. [doi: 10.3233/SHTI190409] [Medline: 31438108]
35. Denecke K, Hochreutener S, Pöpel A, May R. Self-Anamnesis with a Conversational User Interface: Concept and Usability Study. Methods Inf Med 2018 Nov 15;57(5-06):243-252. [doi: 10.1055/s-0038-1675822] [Medline: 30875703]
36. Denecke K, Hochreutener SL, Pöpel A, May R. Talking to Ana: A Mobile Self-Anamnesis Application with Conversational User Interface. 2018 Presented at: International Conference on Digital Health; April 23-26, 2018; Lyon, France. [doi: 10.1145/3194658.3194670]
37. Ireland D, Atay C, Liddle J, Bradford D, Lee H, Rushin O, et al. Hello Harlie: Enabling Speech Monitoring Through Chat-Bot Conversations. Stud Health Technol Inform 2016;227:55-60. [Medline: 27440289]
38. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
39. Oh KJ, Lee D, Ko B, Choi HJ. A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation. 2017 Presented at: 18th IEEE International Conference on Mobile Data Management (MDM); May 29-June 1, 2017; Daejeon, South Korea. [doi: 10.1109/mdm.2017.64]
40. Nikitina S, Callaioli S, Baez M. Smart Conversational Agents for Reminiscence. 2018 Presented at: IEEE/ACM 1st International Workshop on Software Engineering for Cognitive Services (SE4COG); May 28-29, 2018; Gothenburg, Sweden. [doi: 10.1145/3195555.3195567]
41. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
42. Bickmore TW, Schulman D, Sidner CL. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. J Biomed Inform 2011 Apr;44(2):183-197 [FREE Full text] [doi: 10.1016/j.jbi.2010.12.006] [Medline: 21220044]
43. Househ M, Schneider J, Ahmad K, Alam T, Al-Thani D, Siddig MA, et al. An Evolutionary Bootstrapping Development Approach for a Mental Health Conversational Agent. Stud Health Technol Inform 2019 Jul 04;262(3):228-231. [doi: 10.3233/SHTI190060] [Medline: 31349309]
44. Dongkeon L, Oh KJ, Choi HJ. The chatbot feels you - a counseling service using emotional response generation. 2017 Presented at: IEEE International Conference on Big DataSmart Computing (BigComp); February 13-16, 2017; Jeju, South Korea. [doi: 10.1109/bigcomp.2017.7881752]
45. Martínez-Miranda J, Martínez A, Ramos R, Aguilar H, Jiménez L, Arias H, et al. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. J Med Syst 2019 Jun 25;43(8):246. [doi: 10.1007/s10916-019-1387-1] [Medline: 31240494]
46. Lokman AS, Zain JS. An architectural design of Virtual Dietitian (ViDi) for diabetic patients. 2009 Presented at: 2nd IEEE International Conference on Computer Science and Information Technology; August 8-11, 2009; Beijing, China. [doi: 10.1109/iccsit.2009.5234671]
47. Hussain S, Athula G. Extending a Conventional Chatbot Knowledge Base to External Knowledge Source and Introducing User Based Sessions for Diabetes Education. 2018 Presented at: 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA); May 16-18, 2018; Krakow, Poland. [doi: 10.1109/waina.2018.00170]
48. Belfin RV, Shobana AJ, Manilal M, Mathew AA, Babu B. A Graph Based Chatbot for Cancer Patients. 2019 Presented at: 5th International Conference on Advanced Computing Communication Systems (ICACCS; March 15-16, 2019; Coimbatore, India. [doi: 10.1109/icaccs.2019.8728499]
49. Chaix B, Bibault J, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot. JMIR Cancer 2019 May 02;5(1):e12856 [FREE Full text] [doi: 10.2196/12856] [Medline: 31045505]
50. Aljameel SS, O'Shea JD, Crockett KA, Latham A, Kaleem M. Development of an Arabic Conversational Intelligent Tutoring System for Education of children with ASD. 2017 Presented at: IEEE International Conference on Computational Intelligence

and Virtual Environments for Measurement Systems and Applications (CIVEMSA); June 26-28, 2017; Annecy, France. [doi: 10.1109/civemsa.2017.7995296]

51. Vijayan A, Janmasree S, Keerthana C, Syla LB. A Framework for Intelligent Learning Assistant Platform Based on Cognitive Computing for Children with Autism Spectrum Disorder. 2018 Presented at: International CET Conference on Control, Communication, and Computing (IC4); July 5-7, 2018; Thiruvananthapuram, India. [doi: 10.1109/cetic4.2018.8530940]
52. Aarabi P. Virtual cardiologist - A conversational system for medical diagnosis. 2013 Presented at: IEEE Canadian Conference on Electrical and Computer Engineering (CCECE); May 26, 2013; Regina, SK, Canada. [doi: 10.1109/ccece.2013.6567775]
53. Kadariya D, Venkataramanan R, Yip HY, Kalra M, Thirunarayanan K, Sheth A. kBot: Knowledge-Enabled Personalized Chatbot for Asthma Self-Management. 2019 Presented at: IEEE International Conference on Smart Computing (SMARTCOMP); June 12-15, 2019; Washington, DC, USA. [doi: 10.1109/smartcomp.2019.00043]
54. van Heerden A, Ntinga X, Vilakazi K. The potential of conversational agents to provide a rapid HIV counseling and testing services. 2018 Presented at: International Conference on the Frontiers and Advances in Data Science (FADS); October 23-25, 2017; Xi'an, China. [doi: 10.1109/fads.2017.8253198]
55. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The Next Generation: Chatbots in Clinical Psychology and Psychotherapy to Foster Mental Health – A Scoping Review. Verhaltenstherapie 2019 Aug 20:1-13. [doi: 10.1159/000501812]
56. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
57. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
58. Bhirud N, Tataale S, Randive S, Nahar S. A Literature Review On Chatbots In Healthcare Domain. International Journal of Scientific & Technology Research 2019 Jul;8(7):225-231.

Edited by G Eysenbach; submitted 04.04.20; peer-reviewed by K Denecke, J Martínez-Miranda, I Roomaney, JT te Gussinklo; comments to author 12.06.20; revised version received 14.07.20; accepted 20.10.20; published 18.12.20.

Please cite as:
Safi Z, Abd-Alrazaq A, Khalifa M, Househ M
Technical Aspects of Developing Chatbots for Medical Applications: Scoping Review
J Med Internet Res 2020;22(12):e19127
URL: http://www.jmir.org/2020/12/e19127/
doi:10.2196/19127
PMID:33337337

Review

Promotion of Physical Activity in Older People Using mHealth and eHealth Technologies: Rapid Review of Reviews

Lisa McGarrigle[1,2], PhD; Chris Todd[1,2,3,4,5], PhD

[1]School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

[2]Manchester Academic Health Science Centre, Manchester, United Kingdom

[3]National Institute for Health Research Older People and Frailty Policy Research Unit, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

[4]Manchester University NHS Foundation Trust, Manchester, United Kingdom

[5]National Institute for Health Research Applied Research Collaboration-Greater Manchester, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

Corresponding Author:
Chris Todd, PhD
School of Health Sciences
Faculty of Biology, Medicine and Health
The University of Manchester
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 1613067865
Email: chris.todd@manchester.ac.uk

Abstract

Background: Older people are at increased risk of adverse health events because of reduced physical activity. There is concern that activity levels are further reduced in the context of the COVID-19 pandemic, as many older people are practicing physical and social distancing to minimize transmission. Mobile health (mHealth) and eHealth technologies may offer a means by which older people can engage in physical activity while physically distancing.

Objective: The objective of this study was to assess the evidence for mHealth or eHealth technology in the promotion of physical activity among older people aged 50 years or older.

Methods: We conducted a rapid review of reviews using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We searched for systematic reviews published in the English language in 3 electronic databases: MEDLINE, CINAHL Plus, and Scopus. Two reviewers used predefined inclusion criteria to select relevant reviews and extracted data on review characteristics and intervention effectiveness. Two independent raters assessed review quality using the AMSTAR-2 tool.

Results: Titles and abstracts (n=472) were screened, and 14 full-text reviews were assessed for eligibility. Initially, we included 5 reviews but excluded 1 from the narrative as it was judged to be of critically low quality. Three reviews concluded that mHealth or eHealth interventions were effective in increasing physical activity. One review found that the evidence was inconclusive.

Conclusions: There is low to moderate evidence that interventions delivered via mHealth or eHealth approaches may be effective in increasing physical activity in older adults in the short term. Components of successful interventions include self-monitoring, incorporation of theory and behavior change techniques, and social and professional support.

(J Med Internet Res 2020;22(12):e22201) doi:10.2196/22201

KEYWORDS

physical activity; mHealth; eHealth; app; accelerometer; pedometer; technology; COVID-19

Introduction

Older people engage in physical activity less regularly than younger age groups, and participation progressively decreases with age [1]. Inactivity is associated with frailty and adverse health outcomes in middle-aged to older adults [2], with sedentary people aged >50 years having twice the risk of death compared to those with the highest levels of physical activity [3]. The benefits of engagement in physical activity in older age are vast. Regular activity of moderate intensity (150 minutes per week) is consistently associated with reduced risk of chronic diseases [4], cognitive decline [5], and mortality [6]. Exercise programs that emphasize improving strength and balance reduce falls in older people [7].

During 2020, as a result of lockdown and physical or social distancing measures introduced in an effort to reduce coronavirus (COVID-19) transmission, there is concern that older people are at risk of further reduced activity levels and consequently, at increased risk of adverse health events. One approach to promoting activity that has gained particular traction over recent years is the use of exercise and activity mobile apps, tracking devices, and tablet or computer-based interventions. Physical activity interventions delivered via apps on mobile or wireless devices, such as smartphones or tablets, are collectively referred to as mobile health (mHealth) interventions [8]. Interventions delivered or enhanced through the internet and related technologies (eg, websites, wearable motion sensing devices) are referred to as eHealth interventions [9]. These kinds of interventions may offer a cost-effective and accessible way to promote activity in older populations as an alternative to face-to-face sessions. A recent review and meta-analysis of randomized controlled trials (RCTs) across a range of age groups found that mHealth interventions increased physical activity levels and reduced sedentary behavior [10]. mHealth and eHealth approaches may be feasible and acceptable in older populations [11], with an increasing number of older adults accessing the internet in recent years [12]. The effectiveness of these kinds of interventions in older people has been widely investigated, but a clear consensus on their usefulness in increasing physical activity is lacking. The objective of this review was to summarize the evidence from systematic reviews of the effectiveness of mHealth and eHealth approaches on physical activity in older people.

Methods

To provide an overview the evidence on the use of mHealth and eHealth technologies in the promotion of physical activity in older people, we undertook a rapid review of reviews [13,14] and followed guidance for conducting overviews [15]. We followed a standard protocol in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) statement, adjusted for rapid review and "review of reviews" methodologies.

Search Strategy

Database searches were conducted in MEDLINE (Ovid), CINAHL Plus (EBSCO), and Scopus (Elsevier) for reviews focusing on mHealth or eHealth technologies in the promotion of physical activity in older people. All searches were conducted in May 2020 (see Multimedia Appendix 1 for the search strategy).

Inclusion and Exclusion Criteria

As this was a rapid review, included papers were restricted to full-text availability in the English language. No restrictions were made for year of publication nor for study designs included in the reviews in order to get a complete picture of the effectiveness of the interventions.

Reviews were included if they used systematic review techniques to review the use of mHealth or eHealth technologies for the promotion of physical activity among older adults; the mHealth or eHealth interventions targeted generally healthy older people aged ≥50 years; and outcomes were related to levels of physical activity, exercise, fitness, or reduction in sedentary behavior, measured using self-report instruments, measurement devices (eg, accelerometers, pedometers), or energy expenditure (eg, metabolic equivalents of task [METs]).

Reviews were excluded if they did not use systematic review techniques, did not focus on older populations aged ≥50 years, focused on mHealth or eHealth interventions only in disease-specific samples or in people with chronic conditions (eg, Parkinson's disease, stroke, diabetes, cancer, depression, obesity), or focused solely on technology interventions that required equipment other than body-worn sensors (eg, smartwatches) or smartphone-type devices, which are costly and unlikely to be easily accessible to older people (ie, we excluded exergaming, virtual reality gaming, smart homes, robotics).

Reviews in which only a minority of the included studies met the described criteria (ie, only some of the included studies focused on older adults, only some of the included studies focused on generally healthy populations, the majority of technologies were outside our inclusion criteria) were considered for inclusion only if it was agreed between both authors independently that they provided insight into the effectiveness of mHealth or eHealth interventions. In order to be accepted for inclusion, the review had to present analysis for the subset of relevant studies included in the review that fulfilled our inclusion criteria, or it had to be possible to derive estimates of effect for the subset of relevant studies included in the review from data presented.

Screening and Data Extraction

Following identification and removal of duplicates, we exported all citations from Zotero (version 5.0.85) [16] into Rayyan [17], a web application to expedite the blinded screening process. Titles, abstracts, and potentially relevant full texts were screened independently by both authors, and any disagreements were resolved by discussion. Both authors independently used a tailored, predefined, data extraction form to record relevant review characteristics (Multimedia Appendix 2 for data extracted). Descriptions of the included reviews were tabulated for clarity.

Data Analysis

Depending on heterogeneity of interventions and outcome measurements and summary measures used across included systematic reviews, we undertook a meta-analysis [18]. If the meta-analysis was not possible, we present pooled summary data without further analysis.

Quality Assessment

Included reviews were subjected to quality assessment using AMSTAR-2 [19,20]. AMSTAR-2 is used to generate an overall rating of "high," "moderate," "low," or "critically low." Both authors conducted the assessments independently and then discussed ratings to agree consensus. Assessments of study quality or risk of bias reported in the included systematic reviews are presented (Table 1).

Table 1. Summary of review characteristics and quality assessment.

Author (year)	Type of review	Number of included studies and designs	Aim	Population (n)	AMSTAR-2 Rating	Quality of evidence[a]
Cooper et al (2018) [21]	Systematic review and meta-analysis	9 RCTs[b] (8 included in meta-analysis)	To investigate how different wearable activity trackers impact PA[c] levels	Older adults aged ≥65 years (939)	Moderate	Cochrane Risk of Bias Tool: high risk (n=6), unclear risk (n=2), low risk (n=1)
Jonkman et al (2018)[d] [22]	Narrative review, but systematic approach	12 RCTs	To provide an overview of the effectiveness of eHealth interventions in increasing PA in older adults	Community-dwelling adults aged ≥55 years (1208)	Critically low	Not assessed
Larsen et al (2019) [23]	Systematic review and meta-analysis	21 studies: 20 RCTs, 1 RCT with crossover design	To estimate the effect of PAM[e]-based interventions on PA behavior	Adults aged ≥65 years (2783)	High	Cochrane Risk of Bias Tool (v2.0): high risk (n=6), moderate risk (n=10), low risk (n=5)
Muellmann et al (2018) [24]	Systematic review	20 studies: 18 RCTs, 2 quasiexperimental designs	To compare the effectiveness of eHealth interventions promoting PA in older adults	Adults aged ≥55 years (6671)	Moderate	Cochrane Risk of Bias Tool: high risk (n=8), moderate risk (n=11), low risk (n=1)
Yerrakalva et al (2019) [25]	Systematic review and meta-analysis	6 studies (5 included in meta-analysis): 5 RCTs, 1 NRSI[f]	To quantify the effect of mHealth[g] app interventions on sedentary time, PA, and fitness	Community-dwelling adults aged ≥55 years (486)	High	GRADE[h] assessment: moderate-certainty evidence for PA and sedentary time; low-certainty evidence for fitness

[a] Quality of the evidence as assessed by the systematic review authors.

[b] RCTs: randomized controlled trials.

[c] PA: physical activity.

[d] Excluded from this overview based on AMSTAR assessment.

[e] PAM: physical activity monitor.

[f] NRSI: nonrandomized study of interventions.

[g] mHealth: mobile health.

[h] GRADE: Grading of Recommendations, Assessment, Development and Evaluations.

Results

The search process is illustrated in Figure 1. We identified 472 potentially relevant articles, after removal of duplicates. Following title and abstract independent screening, 15 reviews remained for full-text screening by both reviewers. Reasons for exclusion are listed in Multimedia Appendix 3. Overall, 5 reviews were selected for inclusion. Of these, 3 were systematic reviews with meta-analyses that provided pooled estimates of the effect of the technologies reviewed, 1 was a systematic review with a narrative synthesis, and 1 was a narrative review that used systematic review methods (Cochrane guidelines) for search selection and data extraction. Study characteristics and quality assessments for each included review are presented in Table 1. There was 100% agreement on AMSTAR-2 quality ratings without need for consensus discussions. Two of the reviews were rated as high quality and 2 as moderate quality. The quality of 1 review [22] was assessed as critically low, meaning that "the review should not be relied upon to provide an accurate and comprehensive summary of the available evidence" [19]. The review was not strictly a systematic review, although it followed a systematic approach, and this may have contributed to the poor AMSTAR-2 rating. This review is not discussed further. For completeness, details of this review are presented in Tables 1 and 2.

Because of the heterogeneity of methods, interventions, and outcomes, we were unable to undertake a meta-analysis of the reviews included.

Figure 1. PRISMA flow diagram of the inclusion process.

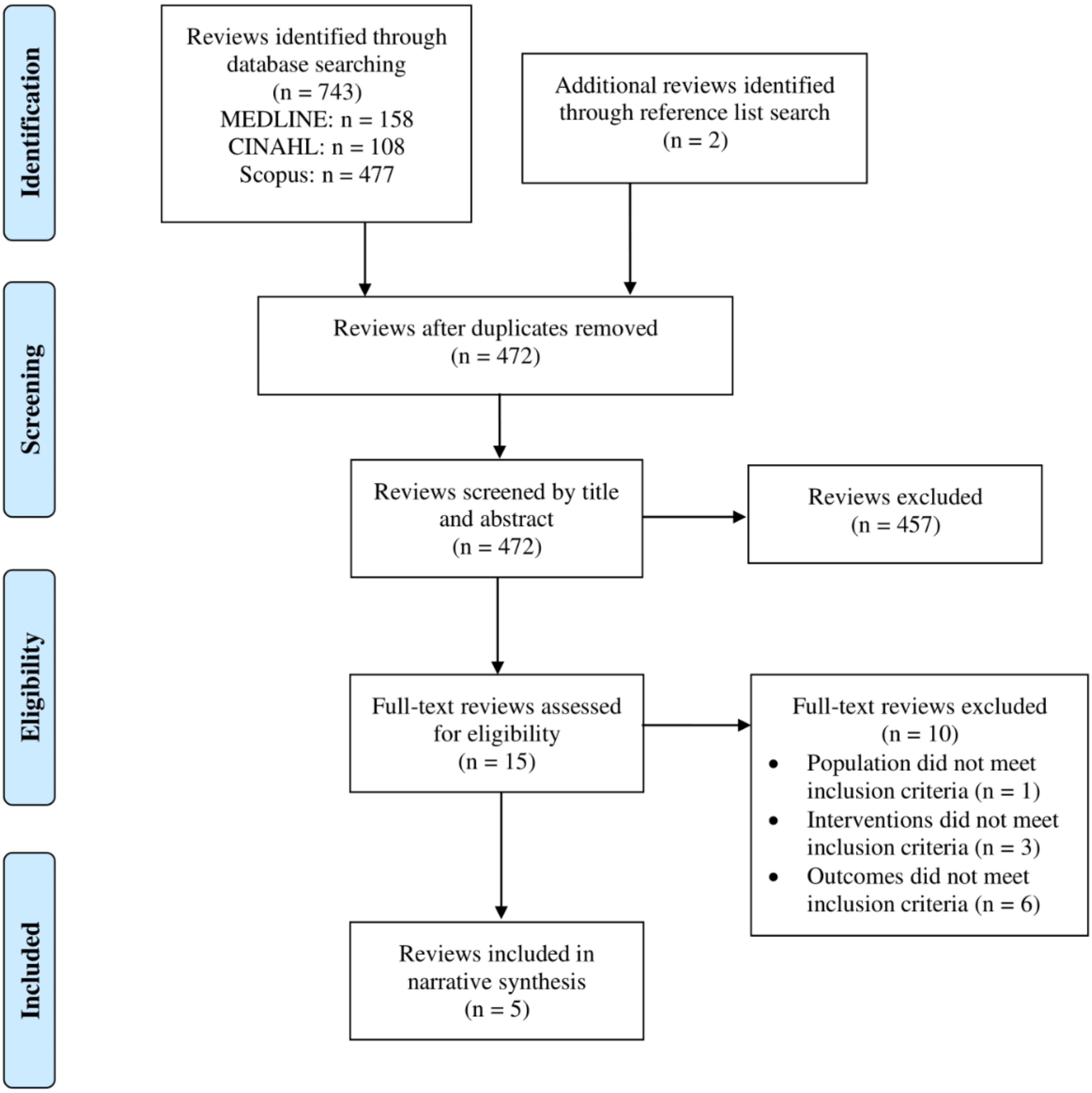


Table 2. Summary of review interventions, comparators, outcomes, and conclusion regarding intervention effectiveness.

Author (year)	mHealth[a] or eHealth intervention	Comparator or control	Key outcomes	Effectiveness of intervention
Cooper et al (2018) [21]	Wearable motion sensing technology interventions (accelerometer or pedometer) designed to increase PA[b]; duration range ≥6 weeks, ≤52 weeks	Usual care or standard care, waitlist control, or other active comparator focused on enhancing PA (eg, educational or behavioral interventions)	Change in PA behavior (minutes walking per day, steps per day, proportion of participants at activity goal)	Statistically significant effect of using accelerometers (SMD[c]=0.43 [95% CI 0.19 to 0.68, I^2=1.6%[d]]), but not pedometers (SMD=0.17 [95% CI −0.08 to 0.43, I^2=37.7%]) for increasing PA levels
Jonkman et al (2018)[e] [22]	Computer, tablet, smartphone, or smartwatch technology to promote PA or reduce sedentary behavior; duration range ≥4 weeks, ≤6 months	Usual practice, waitlist control, or other active comparator (eg, pedometer, blinded activity tracking)	Objective assessment of the amount of PA (eg, daily step counts, minutes spent on PA)	Positive short-term effect of increased PA (ie, right after administering the intervention), but lacking evidence for long-term effects
Larsen et al (2019) [23]	Any PAM[f]-based intervention (ie, accelerometer or pedometer) where the participants of the intervention group received any kind of feedback on their PA level measured by PAMs; duration range ≥4 weeks, ≤52 weeks	No feedback on PA level is given from the PAMs	PA (steps per day) as primary outcome; secondary outcomes included: MVPA[g], sedentary time, physical capacity, HRQoL[h]	Statistically significant effect on PA favoring the intervention (SMD=0.54 [95% CI 0.34 to 0.73, I^2=79.2%]); statistically significant effect on MVPA favoring the intervention (SMD=0.34 [95% CI 0.15 to 0.52, I^2 = 65.8%]); inconclusive results for effects on sedentary time; no effect on physical capacity, BMI, or HRQoL
Muellmann et al (2018) [24]	The main intervention component delivered via computer (ie, website, PDA, virtual advisor), phone, or text messaging; duration range ≥4 weeks, ≤24 months	Non-eHealth intervention (eg, paper-pencil intervention without eHealth component, face-to-face consultation such as prescription of PA by a physician, or exercise in groups or with a personal trainer) or no intervention	PA assessed using objective measures (eg, pedometer, accelerometer), subjective measures (eg, PA diary, questionnaires), or a combination of objective and subjective methods	Overall, the 9 studies that used web-based interventions appear to have beneficial effects on increasing PA compared to various comparators (no intervention or paper-based interventions) in the short-term (1-6 months) with small effect sizes in the range of 0.20-0.31.
Yerrakalva et al (2019) [25]	mHealth app intervention delivered via smartphones or tablet computers; in 5 of 6 studies, the app synced with a wearable device (pedometers or wearable smart device); duration range ≥3 months, ≤6 months	Modified dose of intervention (modified volume of intervention or modified version of same app), different app, non-app intervention, or no intervention	PA (active minutes per day; steps per day), physical fitness (maximal oxygen uptake, 6-minute timed walk, gait speed), and sedentary time (% sedentary time per day, sitting time per day)	Interventions may be associated with increased PA (pooled mean difference 506 steps/day [95% CI −80 to 1092, I^2=80.5%]), decreased sedentary time (SMD=−0.49 [95% CI −1.02 to 0.03, I^2=0%]), and increased fitness (SMD=0.31 [95% CI −0.09 to 0.70, I^2=0%]) in trials ≤3 months and with increased PA (753 steps/day) in trials ≥6 months. Results for all individual outcomes revealed trends in the same direction, but all results were inconclusive as the CIs included zero.

[a] mHealth: mobile health.

[b] PA: physical activity.

[c] SMD: standardized mean difference.

[d] Measure of heterogeneity.

[e] Excluded from this overview based on AMSTAR assessment.

[f] PAM: physical activity monitor.

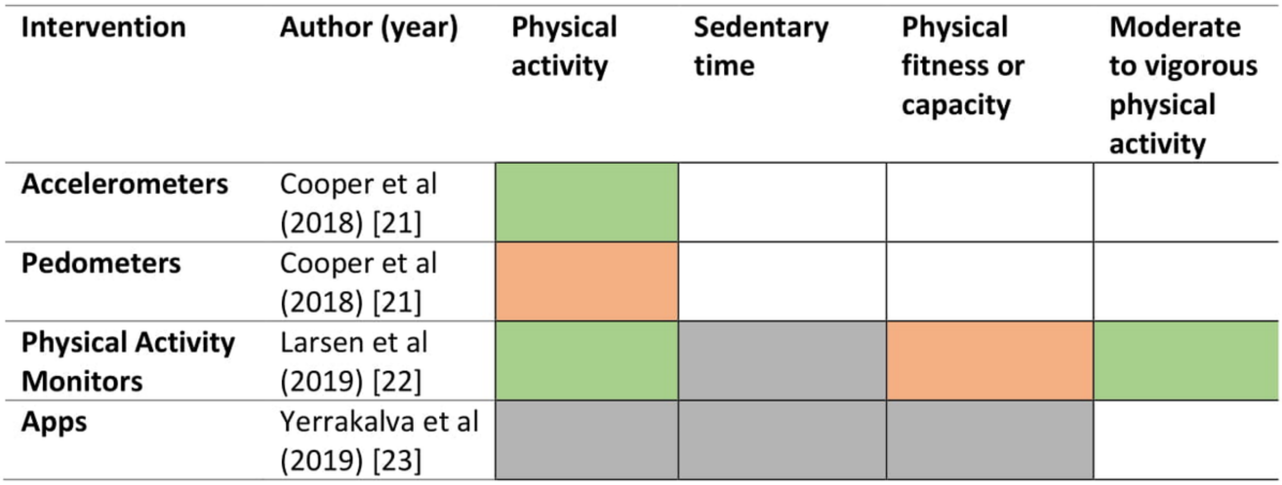
[g] MVPA: moderate-to-vigorous physical activity.

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

Types of mHealth and eHealth Interventions

An overview of the interventions, comparators, outcomes, and effectiveness of the interventions described in the reviews are presented in Table 2. In 2 reviews, interventions were delivered via wearable motion sensing devices (eg, accelerometers or pedometers) [21,23]. In 1 review, interventions were delivered via apps on smartphones or tablets, and all but one of the included studies involved syncing the app to wearable devices [25]. One review [24] included a mix of interventions, including computer-based (ie, websites, personal digital assistant, virtual advisor; n=9), phone (n=7), and text messaging (n=4), with a separate narrative synthesis provided for each mode of intervention. The overall intervention duration for studies included in reviews ranged from 4 weeks to 2 years. All reviews reported on physical activity as a key outcome. Figure 2 presents a summary of evidence from meta-analyses for each intervention.

Figure 2. Summary of evidence from meta-analyses. Green indicates the intervention was effective, amber indicates there was no difference in the investigated comparison, grey indicates the evidence was inconclusive, and no color (blank) indicates the outcome was not assessed.



Intervention	Author (year)	Physical activity	Sedentary time	Physical fitness or capacity	Moderate to vigorous physical activity
Accelerometers	Cooper et al (2018) [21]	🟩			
Pedometers	Cooper et al (2018) [21]	🟧			
Physical Activity Monitors	Larsen et al (2019) [22]	🟩	▨	🟧	🟩
Apps	Yerrakalva et al (2019) [23]	▨	▨	▨	

Reviews Reporting Significant Effects of Interventions

Significant effects of mHealth or eHealth interventions on physical activity outcomes were reported by 3 reviews.

Larsen et al [23] reviewed 21 studies on the effects of physical activity monitors (PAM) on physical activity, sedentariness, physical capacity, BMI, and self-reported health-related quality of life (HRQoL) compared to control interventions. Of these studies, 20 were RCTs with parallel group design, and 1 was a crossover RCT. Although 6 studies included samples with specific diagnoses, subgroup analysis revealed no significant differences across any of the outcomes in relation to diagnoses. Regarding the risk of bias, 5 studies were rated as having low risk of bias, 10 studies as medium risk, and 6 studies as high risk of bias. Results of a random effects meta-analysis on the effect of PAM intervention on physical activity found a moderate standardized mean difference (SMD) of 0.54 (95% CI 0.34 to 0.73) in support of the PAM interventions, equivalent to about 1300 more steps per day. Furthermore, there was a small to moderate effect on moderate-to-vigorous physical activity equivalent to 8 more minutes activity per day (SMD 0.35, 95% CI 0.15 to 0.52), favoring intervention. Heterogeneity was high for both comparisons. The pooled effects for time spent sedentary, physical capacity, BMI, and self-reported HRQoL were not significant. Effects are not reported in relation to length of the intervention, but sensitivity analysis found that intervention length was not significantly correlated with the effect size for any outcomes. Overall, the quality of the evidence was judged as low to moderate due to unexplained heterogeneity, publication bias, and imprecision. However, the authors conclude that PAM-based interventions are safe and feasible for use in older adult populations.

Cooper et al [21] reviewed 9 RCTs that investigated the effect of accelerometer and pedometer use on physical activity in older adults, 8 of which were included in the meta-analysis. Most studies (n=6) were judged to be at high risk of bias, 2 were unclear, and 1 was considered to be at low risk of bias. Pooled results from 4 studies investigating the effect of accelerometers revealed small to moderate positive effects on physical activity (SMD 0.43, 95% CI 0.19 to 0.68). Pooled results from 4 studies investigating pedometers revealed no statistically significant effect on physical activity (SMD 0.22, 95% CI –0.08 to 0.51). Overall, shorter duration accelerometer interventions appeared to have a larger effect, but intervention duration and individual study estimates were variable. Adherence to the intervention was reported in 5 of the 9 included studies and was high on average (≥80%). The authors concluded that higher step detection accuracy in accelerometers may explain why feedback from accelerometers was found to increase levels of physical activity, whereas pedometers did not, but caution that the high risk of bias found in most studies limits conclusions that can be drawn from these findings.

Muellmann et al [24] reviewed 20 studies comparing the effects of eHealth interventions on physical activity in older people (≥55 years). Only 1 study was assessed as having low risk of bias. Heterogeneity in intervention, mode of delivery, duration, outcome assessments, and comparator groups precluded meta-analysis. There were 18 RCTs and 2 quasiexperiments. Website-based interventions were used in 9 studies. These studies overall appear to have beneficial effects on increasing physical activity compared to comparators (no intervention or paper-based interventions) in the short-term (1-6 months) with small effect sizes in the range of 0.20-0.31. Of the 7 studies using telephone-based interventions (health education, telephone fitness sessions, and advice), 3 studies reported no effect, while 4 reported effects of improved activity. Observed effects were most common over the shorter term. Of the 4 studies including text-messaging interventions, 3 reported positive effects of text messaging over periods up to 3 months but not beyond. The authors conclude that "eHealth interventions can effectively promote PA in older adults aged 55 years and above when compared to no intervention control groups at least in the short term." Interventions that are theory-based were more effective regardless of intervention mode, and greater engagement was associated with effect, but participants seldom reached intended exercise dose. The authors urge caution, as the risk of bias in the studies reviewed was high to moderate and there was great heterogeneity in intervention mode, content, and duration.

Reviews Reporting Inconclusive Effects of Interventions

One of the reviews, by Yerrakalva et al [25], found no significant effects of app-based interventions on physical activity outcomes. The review included 6 studies (5 RCTs); studies appear underpowered, with all but 1 having total sample sizes <65 and intervention groups <25 participants. For the RCTs, risk of bias was judged to be low. The pooled analyses reported the following: for physical activity, average increases of 506 steps per day over 3 months and 753 steps per day over 6-12 months; for sedentary behavior, reductions in sedentary time at 3 months and 6 months; for physical fitness, small increases in fitness up to 3 months expressed in a number of different ways (gait speed, timed walk, VO_2 max). Results reveal trends in the same direction for an effect of the apps but have to be judged as inconclusive as confidence intervals included zero. Features that appeared to be common to apps demonstrating improvement trends included syncing to smartwatches or activity monitors and behavior change techniques including goal setting, self-monitoring, instructions for use, social rewards, and combining the app with professional support.

Discussion

We undertook a rapid review of systematic reviews to assess the evidence for mHealth or eHealth interventions in the promotion of physical activity among older people. We initially included 5 reviews but excluded 1 (critically low quality) from the overview. We were unable to undertake meta-analysis of the reviews because of the heterogeneity of the methods, interventions, and outcomes. Three reviews [21,23,24], 2 of which included meta-analyses, found that eHealth approaches (activity monitors, web-based interventions) improved physical activity outcomes. One review and meta-analysis [25] reported trends toward an intervention effect; however, results were inconclusive. Although risk of bias was judged to be low in the studies included in the review, they appeared underpowered, with small total and group sample sizes, potentially explaining why no significant effects were found. Overall, there is evidence from 3 moderate- to high-quality systematic reviews to support the effectiveness of mHealth or eHealth approaches in increasing physical activity, at least over the short term. Due to reported risk of bias in the studies included in the reviews, the overall quality of this evidence is judged as low to moderate.

The reviews identified several components of successful interventions that align with behavior change techniques known to increase motivation and facilitate behavior modification [26]. Most of the studies involved elements of self-monitoring and feedback, either in the form of apps or wearable devices. The provision of real-time feedback on individuals' levels of physical activity was associated with significant increases in the behavior [24], and the implementation of motivational tools such as self-monitoring and feedback as a means of positively impacting levels of physical activity, goal attainment, and adherence has previously demonstrated success [21]. Interventions that were theory-based were more effective, and effect was associated with greater engagement regardless of the mode of intervention [24]. Although Yerrakalva et al [25] determined the effect of the intervention was inconclusive, they reported on the common features of the studies that found app interventions to be effective. These included self-monitoring, goal setting, clear instruction on how to perform the behavior, and social and professional support. Support from professionals may be particularly important in encouraging engagement and adherence to home exercise. A review on home-based, nontechnology physical activity interventions in older adults [27] found that contacting individuals by phone in order to provide support was a good alternative to onsite supervision. They reported strong evidence, based on 3 studies, indicating that direct remote contact had positive effects on physical activity and capacity measures, to a similar extent as supervised training.

An important consideration for mHealth or eHealth interventions concerns the practicalities of implementing technological solutions, especially when users have limited experience of using such devices [28]. This is an important point to keep in mind for the older and disadvantaged sectors of the population, whom we already know are likely to have less experience of technologies [29]. Consideration must be given to the "digital divide" that may result from socioeconomic status, age, geographic location, and cultural factors [28]. Technologies need to provide activity interventions that fit in with older peoples' lifestyles and expectations and offer tailored interventions taking account of individual preferences and capabilities. A review of older adults' perceptions of technologies [30] found that one size does not fit all and technologies need to be tailored to individual need. Intrinsic factors related to older adults' attitudes towards technologies centered around control, independence, and perceived need or requirement for safety are all important for motivation to use the technologies. Extrinsic factors identified were related to usability, feedback gained from technology, and costs. If older people are to use technologies, the positive benefits need to be emphasized and clearly recognizable, including how the technology promotes independence. Technologies need to be perceived as reliable and effective if they are to engender long-term use. Acceptability of technology and adherence to the interventions were not key outcomes in any of the reviews included in our review. Even so, adherence was high on average in 5 of the studies included in 1 review [21]. Wearable devices such as accelerometers and pedometers are feasible for use with older adult populations, with devices attached to the wrist or ankle being most commonly accepted in older groups [31].

This rapid review of reviews used rigorous systematic methods in the search of the literature and the assessment of review quality, and we excluded reviews of low methodological quality. Nevertheless, reviews of reviews can be limited as they face the challenge of synthesizing information in the presence of inevitable heterogeneity. This had a clear effect in this review resulting in our inability to meta-analyze results from previous reviews. Even so, this type of review allows a greater scope for generality in research findings and provides an accessible summary of evidence to support decision making by health care professionals or policy makers [14].

In conclusion, the use of mHealth or eHealth interventions with older people may be effective in increasing physical activity and physical fitness and decreasing sedentary time, over the short term. However, the evidence is currently not conclusive. These findings are in line with those relating to mHealth interventions to increase activity and reduce sedentary behaviors for all age groups [10]. Furthermore, our review suggests that interventions that are theory-based and include behavior change techniques, clear instructions, and social and professional support may be more effective than those that do not. When introducing new technologies such as apps to older people, the steep learning curve older people may experience must be recognized and support supplied to help them become familiar with the technology. Given the ongoing crisis caused by COVID-19 and the challenges we face as we reorganize services in a post-COVID-19 world, there is clearly great potential for digitalization of services for older people, although differences in uptake could exacerbate health inequalities if access is not made available to all groups. The evidence for the effectiveness of mHealth or eHealth provision of interventions to support and promote physical activity among older people is still in its infancy, but nonetheless promising. Future research requires high-quality RCTs comparing different modes of delivery, but implementation may require faster turnaround and should therefore be rigorously evaluated.

Acknowledgments

This work was supported by the National Institute for Health Research through the NIHR Policy Research Unit in Older People and Frailty, NIHR Applied Research Collaboration Greater Manchester, and a NIHR Senior Investigator Award to CT. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care, or its partner organizations.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms.
[PDF File (Adobe PDF File), 100 KB - jmir_v22i12e22201_app1.pdf]

Multimedia Appendix 2
Data extracted.
[PDF File (Adobe PDF File), 59 KB - jmir_v22i12e22201_app2.pdf]

Multimedia Appendix 3
Excluded full text articles with reasons.
[PDF File (Adobe PDF File), 109 KB - jmir_v22i12e22201_app3.pdf]

References

1. Sun F, Norman IJ, While AE. Physical activity in older people: a systematic review. BMC Public Health 2013 May 06;13(1):449 [FREE Full text] [doi: 10.1186/1471-2458-13-449] [Medline: 23648225]
2. Blodgett J, Theou O, Kirkland S, Andreou P, Rockwood K. The association between sedentary behaviour, moderate-vigorous physical activity and frailty in NHANES cohorts. Maturitas 2015 Feb;80(2):187-191. [doi: 10.1016/j.maturitas.2014.11.010] [Medline: 25542406]
3. Banks J, Breeze E, Lessof C, Nazroo J. Mortality and healthy life expectancy. Living in the 21st Century: Older People in England.: The Institute for Fiscal Studies; 2008. URL: https://www.ifs.org.uk/elsa/report08/elsa_w3.pdf [accessed 2020-12-12]
4. Blair SN, Morris JN. Healthy hearts - and the universal benefits of being physically active: physical activity and health. Ann Epidemiol 2009 Apr;19(4):253-256. [doi: 10.1016/j.annepidem.2009.01.019] [Medline: 19344864]
5. Koščak Tivadar B. Physical activity improves cognition: possible explanations. Biogerontology 2017 Aug 10;18(4):477-483. [doi: 10.1007/s10522-017-9708-6] [Medline: 28492999]

6. Arem H, Moore SC, Patel A, Hartge P, Berrington de Gonzalez A, Visvanathan K, et al. Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. JAMA Intern Med 2015 Jun 01;175(6):959-967 [FREE Full text] [doi: 10.1001/jamainternmed.2015.0533] [Medline: 25844730]
7. Sherrington C, Fairhall N, Wallbank G, Tiedemann A, Michaleff ZA, Howard K, et al. Exercise for preventing falls in older people living in the community. Cochrane Database Syst Rev 2019 Jan 31;1:CD012424 [FREE Full text] [doi: 10.1002/14651858.CD012424.pub2] [Medline: 30703272]
8. Bateman DR, Srinivas B, Emmett TW, Schleyer TK, Holden RJ, Hendrie HC, et al. Categorizing Health Outcomes and Efficacy of mHealth Apps for Persons With Cognitive Impairment: A Systematic Review. J Med Internet Res 2017 Aug 30;19(8):e301 [FREE Full text] [doi: 10.2196/jmir.7814] [Medline: 28855146]
9. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
10. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth Technologies to Influence Physical Activity and Sedentary Behaviors: Behavior Change Techniques, Systematic Review and Meta-Analysis of Randomized Controlled Trials. Ann Behav Med 2017 Apr 18;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
11. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and Acceptability of a Wearable Technology Physical Activity Intervention With Telephone Counseling for Mid-Aged and Older Adults: A Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
12. Gell N, Rosenberg D, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421 [FREE Full text] [doi: 10.1093/geront/gnt166] [Medline: 24379019]
13. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012 Feb 10;1(1):10 [FREE Full text] [doi: 10.1186/2046-4053-1-10] [Medline: 22587960]
14. Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: planning a relevant research question and objective for an overview. Syst Rev 2018 Mar 01;7(1):39 [FREE Full text] [doi: 10.1186/s13643-018-0695-8] [Medline: 29490699]
15. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011 Feb 03;11(1):15 [FREE Full text] [doi: 10.1186/1471-2288-11-15] [Medline: 21291558]
16. Roy Rosenzweig Center for History and New Media. Zotero. 2020. URL: https://www.zotero.org/ [accessed 2020-12-14]
17. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
18. Pollock M, Fernandes R, Becker L. Chapter V: Overviews of Reviews. In: Higgins J, Thomas J, Chandler J, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 6.0 (Updated March 2020). London: Cochrane; 2020.
19. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
20. AMSTAR. 2017. URL: www.amstar.ca [accessed 2020-12-12]
21. Cooper C, Gross A, Brinkman C, Pope R, Allen K, Hastings S, et al. The impact of wearable motion sensing technology on physical activity in older adults. Exp Gerontol 2018 Oct 02;112:9-19 [FREE Full text] [doi: 10.1016/j.exger.2018.08.002] [Medline: 30103025]
22. Jonkman NH, van Schooten KS, Maier AB, Pijnappels M. eHealth interventions to promote objectively measured physical activity in community-dwelling older people. Maturitas 2018 Jul;113:32-39. [doi: 10.1016/j.maturitas.2018.04.010] [Medline: 29903646]
23. Larsen RT, Christensen J, Juhl CB, Andersen HB, Langberg H. Physical activity monitors to enhance amount of physical activity in older adults - a systematic review and meta-analysis. Eur Rev Aging Phys Act 2019 May 4;16(1):7 [FREE Full text] [doi: 10.1186/s11556-019-0213-6] [Medline: 31073341]
24. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: A systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
25. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of Mobile Health App Interventions on Sedentary Time, Physical Activity, and Fitness in Older Adults: Systematic Review and Meta-Analysis. J Med Internet Res 2019 Nov 28;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
26. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug 20;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
27. Geraedts H, Zijlstra A, Bulstra SK, Stevens M, Zijlstra W. Effects of remote feedback in home-based physical activity interventions for older adults: a systematic review. Patient Educ Couns 2013 Apr;91(1):14-24. [doi: 10.1016/j.pec.2012.10.018] [Medline: 23194823]

28. Carter DD, Robinson K, Forbes J, Hayes S. Experiences of mobile health in promoting physical activity: A qualitative systematic review and meta-ethnography. PLoS One 2018 Dec 17;13(12):e0208759 [FREE Full text] [doi: 10.1371/journal.pone.0208759] [Medline: 30557396]
29. Matthews K, Nazroo J, Marshall A. Digital inclusion in later life: cohort changes in internet use over a ten-year period in England. Ageing and Society 2018 Aug 02;39(9):1914-1932. [doi: 10.1017/s0144686x18000326]
30. Hawley-Hague H, Boulton E, Hall A, Pfeiffer K, Todd C. Older adults' perceptions of technologies aimed at falls prevention, detection or monitoring: a systematic review. Int J Med Inform 2014 Jun;83(6):416-426. [doi: 10.1016/j.ijmedinf.2014.03.002] [Medline: 24798946]
31. de Bruin ED, Hartmann A, Uebelhart D, Murer K, Zijlstra W. Wearable systems for monitoring mobility-related activities in older people: a systematic review. Clin Rehabil 2008 Oct;22(10-11):878-895. [doi: 10.1177/0269215508090675] [Medline: 18955420]

Abbreviations

HRQoL: health-related quality of life
METs: metabolic equivalents of task
PAM: physical activity monitor
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by G Eysenbach; submitted 06.07.20; peer-reviewed by J Alvarez Pitti, AS Aslam; comments to author 04.09.20; revised version received 18.09.20; accepted 29.10.20; published 29.12.20.

Please cite as:
McGarrigle L, Todd C
Promotion of Physical Activity in Older People Using mHealth and eHealth Technologies: Rapid Review of Reviews
J Med Internet Res 2020;22(12):e22201
URL: http://www.jmir.org/2020/12/e22201/
doi:10.2196/22201
PMID:33372894

Review

The Online Health Information Needs of Family Physicians: Systematic Review of Qualitative and Quantitative Studies

Piet van der Keylen[1], BSc, MSc, Dr rer biol hum; Johanna Tomandl[1], BA, MA; Katharina Wollmann[2,3], BA, MPH; Ralph Möhler[4], MSCN, Dr rer med; Mario Sofroniou[5], BSc (Hons), MB BCh (Wales), MRCGP, MMus, DFSRH, Dr; Andy Maun[5], Prof Dr; Sebastian Voigt-Radloff[3], MSc, PhD, Dr; Luca Frank[1], Dr med

[1]Friedrich-Alexander University Erlangen-Nürnberg, Institute of General Practice, University Hospital Erlangen, Erlangen, Germany

[2]Institute for Evidence in Medicine, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

[3]Cochrane Germany Foundation, Freiburg, Germany

[4]Institute for Health Services Research and Health Economics, Heinrich-Heine-University Düsseldorf, Center for Health and Society, Faculty of Medicine, Düsseldorf, Germany

[5]Division of General Practice, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Corresponding Author:
Piet van der Keylen, BSc, MSc, Dr rer biol hum
Friedrich-Alexander University Erlangen-Nürnberg
Institute of General Practice
University Hospital Erlangen
Universitätsstr 29
Erlangen, 91054
Germany
Phone: 49 91318544953
Email: Piet.Keylen@uk-erlangen.de

Abstract

Background: Digitalization and the increasing availability of online information have changed the way in which information is searched for and retrieved by the public and by health professionals. The technical developments in the last two decades have transformed the methods of information retrieval. Although systematic evidence exists on the general information needs of specialists, and in particular, family physicians (FPs), there have been no recent systematic reviews to specifically address the needs of FPs and any barriers that may exist to accessing online health information.

Objective: This review aims to provide an up-to-date perspective on the needs of FPs in searching, retrieving, and using online information.

Methods: This systematic review of qualitative and quantitative studies searched a multitude of databases spanning the years 2000 to 2020 (search date January 2020). Studies that analyzed the online information needs of FPs, any barriers to the accessibility of information, and their information-seeking behaviors were included. Two researchers independently scrutinized titles and abstracts, analyzing full-text papers for their eligibility, the studies therein, and the data obtained from them.

Results: The initial search yielded 4541 studies for initial title and abstract screening. Of the 144 studies that were found to be eligible for full-text screening, 41 were finally included. A total of 20 themes were developed and summarized into 5 main categories: *individual needs* of FPs before the search; *access needs*, including factors that would facilitate or hinder information retrieval; *quality needs* of the information to hand; *utilization needs* of the information available; and *implication needs* for everyday practice.

Conclusions: This review suggests that searching, accessing, and using online information, as well as any pre-existing needs, barriers, or demands, should not be perceived as separate entities but rather be regarded as a sequential process. Apart from accessing information and evaluating its quality, FPs expressed concerns regarding the applicability of this information to their everyday practice and its subsequent relevance to patient care. Future online information resources should cater to the needs of the primary care setting and seek to address the way in which such resources may be adapted to these specific requirements.

(J Med Internet Res 2020;22(12):e18816) doi:10.2196/18816

KEYWORDS

family physicians; general practitioners; primary care; needs; barriers; online information; health information; health resources; internet; information-seeking behaviors; mobile phone

Introduction

Background

Digital transformation and the ubiquitous availability of online information resources have diversified the process of obtaining and evaluating information in general. Although information availability has seen a transition from classical sources of information to digital equivalents, it has brought with it new barriers such as technical requirements, paying systems, or the need for paid membership to access certain contents. Following initial work on the information needs of doctors by Covell et al [1], reviews have summarized the needs, information-seeking behaviors, and resources used to answer clinical questions that have arisen from everyday practice [2-12]. One reason for clinicians to conduct an information search is to answer questions arising from their daily practice and patient care [12]. Doctors therefore frequently use the internet for professional purposes [13]. They encounter evermore internet-informed patients, who bring information into the consultation [14] and use the internet as their preferred source of health information [15,16]. This inevitably affects the doctor-patient interaction [17] and health-related decision making [18]. The variety and abundance of online medical information may be overwhelming when it comes to critically appraising and evaluating the quality of these resources [19]. Doctors thus face new challenges when it comes to the utilization and adoption emerging through digital transformation in health care. The following questions should be raised: What reasons, facilitators, and barriers exist for doctors during online information searches? How are their information needs and information-seeking behaviors may be affected by digital transformation? Despite the body of available literature, we identified 3 gaps in the literature on the information needs among doctors, leading us to conduct this systematic review:

1. Not family physician (FP)–specific: Half of the systematic reviews examined were not specific to primary care or FP, but on the information needs of doctors in general [4,7,9,10,12], whereas reviews addressing the information needs and information-seeking behaviors of FPs are outdated [2,3,5,6,8]. The latest review by Clarke et al [11] in 2013, analyzed the information-seeking behavior of FPs, trainees, nurse practitioners, nurses, and nurse coordinators in a combined review to better understand clinical decision making. As the daily routines of medical specialties differ greatly, so do the respective information needs of doctors [9], which arise from tackling specific clinical tasks in everyday practice [20]. FPs are confronted with diverse clinical questions and therefore may have differing information requirements than other specialist colleagues. These requirements could be met with a multitude of available online tools and software systems developed within the recent decade. To the best of our knowledge, no recent systematic review has exclusively examined the information needs of FPs toward online health information.
2. Out of date: One review from 2011 studied information needs and information-seeking behaviors of hospital-based doctors compared with primary care physicians regarding the access of electronic information [21]. However, technological advancements as well as rapidly changing information delivery systems over the last decade have altered information retrieval in general and specifically in the health care setting.
3. Contradictory evidence: It remains questionable whether the perceived needs of doctors reflect their actual needs [9]. Existing, but not perceived, needs of physicians could remain unexpressed. The analysis of barriers and facilitators before, during, and after the information search itself may give insights into existing but unperceived or unexpressed needs [9].

Objectives

This systematic review asks “What needs, demands, barriers and facilitators exist for FPs to search for online health information?” We intend to fill these gaps in the literature and aim to do the following:

1. Review studies that analyze the information needs and information-seeking behaviors of FPs in the primary care setting.
2. Focus on online information retrieval by considering the technological advancements in health care and medical information over the last 20 years.
3. Include factors that facilitate or hinder the need for and retrieval of online information in the FP setting.

Therefore, this study intends to summarize the 3 elements of information need, literature searches, and resources as they are interlinked, as suggested by Davies [9].

Methods

Methodological Approach

We performed a systematic review.

Search Strategy

We searched for relevant studies using MEDLINE (via PubMed), Web of Science Core Collection (SCI-Expanded), and Scopus. Furthermore, the reference lists of identified primary papers were screened to identify other potentially relevant citations. The initial search in all databases was performed on May 2, 2018, and updated on January 21, 2020. We formed a search strategy in cooperation with Cochrane Germany and a consulting medical colleague. We started an explorative search term comprising “physician* AND health information AND need.” For the specifications of the explorative search strategy, we searched for relevant synonyms and corresponding Medical Subject Headings (MeSH) terms to extend the explorative terms. These were matched with *MeSH On Demand* and with MeSH terms in similar papers, retrieved by using the explorative search. We established generated blocks for each aspect of the review question, such as the participants involved, the areas of interest, and the setting. Synonyms or similar MeSH terms within each block were combined with the OR operator. The blocks were then combined with the AND operator. We limited our search to the years spanning 2000 to 2020 for the following reasons: global internet access was only widely available from the year 2000 onwards. Subsequently, increased use of the internet could be assumed. This period covers milestone technical developments, such as broadband and mobile internet access, smartphone development and the accompanying hardware and software changes, and social media utilization. Before 2000, only about 5% [22] of the world population had internet access. Thus, it seems reasonable to limit the timeframe accordingly. The final search terms and search details used are provided in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

We included original qualitative, quantitative, and mixed methods studies, which assessed the needs of FPs and their requirements for online health information, regardless of the medical indication. These studies included those that assessed these needs explicitly or more implicitly measured requirements, barriers, and demands in asking clinical questions during an FPs' working day or during continuous medical education (CME) programs. If studies addressed a variety of professions or specialties, we only considered those that consisted of at least 50% FPs in the study population. We included studies retrieved by the search in German and English only, regardless of the impact factor, peer-review process, or publishing process (eg, book, journal, and dissertation).

We excluded the following:

- Reviews, conference proceedings, evidence syntheses, editorials, commentaries, study register entries, protocols, or works that were unobtainable.
- Studies conducted in developing countries that had a very different or underdeveloped health care system, if a reasonable comparison with the primary care systems of the included studies (eg, Germany, the United Kingdom, and France) was not feasible.
- Double publications that only received minor edits or updates to the initial study, by choosing those with the most complete data set.

Furthermore, studies were deemed to be ineligible if:

- The wrong type of information was addressed: As this review aimed at the needs of FPs in obtaining online health information, we excluded:
    - Studies that only analyzed health information in exclusively printable (nononline) formats.
    - Studies that focused on electronic health records or systems aiming at patient information.
- The wrong population was addressed: patients, general public, or nonmedical health professionals (eg, studies that addressed evaluation or utilization of patient-centered telephone or online counseling, interactive apps, online forums, social media, or patient portals with protected log-ins or personalized patient data).
- Only outcomes not connected to information-seeking behavior or information needs toward online health information were measured (eg, piloting and evaluation of specific knowledge interventions or educational programs for physicians).

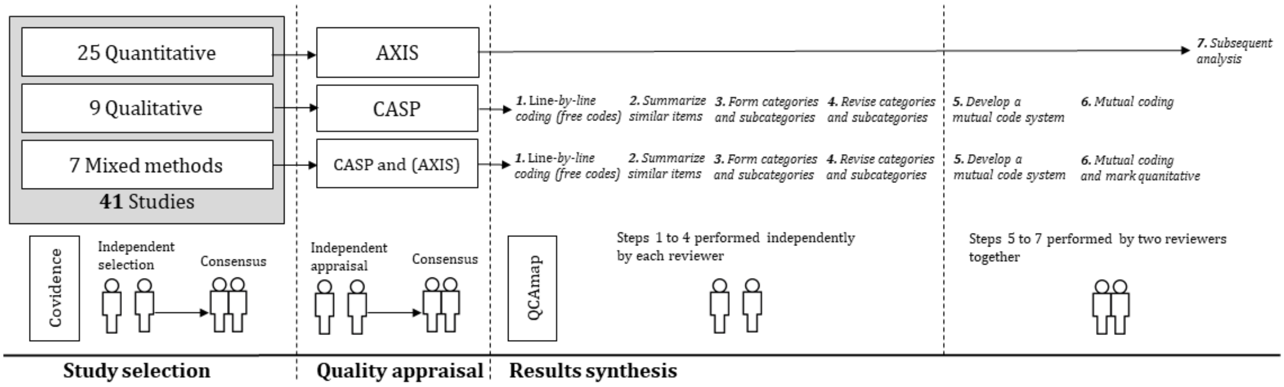
Study Selection

All duplicates were removed automatically using Endnote X9.3.3 (Clarivate) and subsequently by hand. Study selection was performed with Covidence [23], which is specifically designed for Cochrane reviews and frequently used for review management [24,25]. Two researchers (LF and PK) independently screened titles and abstracts and excluded studies that were not eligible. LF and PK then independently screened full-text copies of potentially relevant papers, excluding any studies that were not eligible and documenting the reasons for exclusion. Any disagreements in any phase were resolved through discussion and consensus. Due to the broad themes of independently screened papers, we did not measure inter-rater reliability (kappa) because a quantitative measurement of agreement would not have reflected the qualitative consensus process.

Data Extraction

Two reviewers developed and piloted a data extraction form independently and manually and obtained the following study characteristics: authors, publication date, title, study type (quantitative, qualitative, or mixed methods), type of data collection (questionnaires, interviews, etc), recruitment of participants (email, hospital, etc), number of FP participants, indication for the health information addressed, the type of health information (online, app, etc), and the outcome variables (needs and requirements). The reviewers resolved any disagreements in the extraction phase through discussion and consensus. An overview of the data extraction phase, including quality appraisal, results analysis, and synthesis, is displayed in Figure 1.

Figure 1. Methods overview for study selection, quality appraisal, and results synthesis (for details see text). AXIS: appraisal tool for cross-sectional studies; CASP: critical appraisal skills program.



Quality Appraisal of the Studies

The quality of the studies was assessed by using 2 instruments. For qualitative studies, we applied the critical appraisal skills program (CASP) checklist [26]. For quantitative studies, we applied the appraisal tool for cross-sectional studies (AXIS) tool [27]. For mixed methods studies, we applied the CASP checklist and, if applicable, the AXIS tool. CASP offers distinct and easy-to-use checklists for critically appraising different forms of evidence and studies adapted from the *JAMA Users Guide to the Medical Literature* [28] and is used in similar reviews [29]. The AXIS tool is a new appraisal tool for cross-sectional studies, currently cited in more than 60 reviews as well as in a recent systematic review assessing the effectiveness of apps [30]. As many of the included quantitative studies had a cross-sectional design, it seemed suitable for this review. AXIS offered no numerical scale to assess quality, but instead aimed to assess the individual characteristics of a study and therefore seemed suitable, as quantitative data were too heterogeneous to perform meta-analyses. Critical appraisal also served as an indicator of possible strengths and weaknesses of the studies and any possible implications for thematic synthesis. All quality aspects were extracted independently by 2 reviewers (LF and PK) in a separate data sheet and were later combined by discussion and consensus into the tables shown in Multimedia Appendices 2 [31-46] and 3 [21,32,34,36,40,47-70].

Analysis and Synthesis of Results

A structured analysis of all included studies as well as thematic synthesis was performed based on the method described by Thomas and Harden [71] and according to a systematic review by Möhler and Meyer [72]. All full texts, figures, tables, and supplementary materials of the included studies were uploaded to a qualitative content analysis software (QCAmap), according to Mayring [73]. In studies where mixed populations were addressed, but FPs made up at least 50% of the study population, we only extracted qualitative and quantitative data if they were represented separately. They were excluded from the study if data regarding FPs were not separately extracted. Data synthesis was performed in 7 stages (Figure 1):

1. The results of the qualitative and mixed methods studies were coded line by line according to the meaning of the content (free codes). The free codes were named as per *the description of the item*, enabling the reader to identify the way in which an item was defined in the text of the included studies.
2. A superior item name was generated to summarize the descriptions of similar items.
3. Item names were organized into related areas of descriptive themes (main categories and subcategories) by inductive category formation [73].
4. These categories were compared within the studies to analyze similarities or differences, or create new categories, if existing ones were insufficient. Stages 1 to 4 were performed by both reviewers (LF and PK) independently, thereby generating 2 separate code systems.
5. Both reviewers compared and discussed their respective code systems to achieve consensus on a mutual code system.
6. With this mutual code system, the 2 reviewers together again coded all full-text papers line by line, resulting in a more general and objective coding system during text interpretation. In mixed methods studies, where quantitative content showed relevant connections to qualitative content, this content was additionally marked with a quantitative code to later allow linking to both qualitative and quantitative data sets.
7. Purely quantitative studies were subsequently analyzed to support the main categories and subcategories derived from qualitative synthesis in a sequential synthesis design [74] and without conducting meta-analyses.

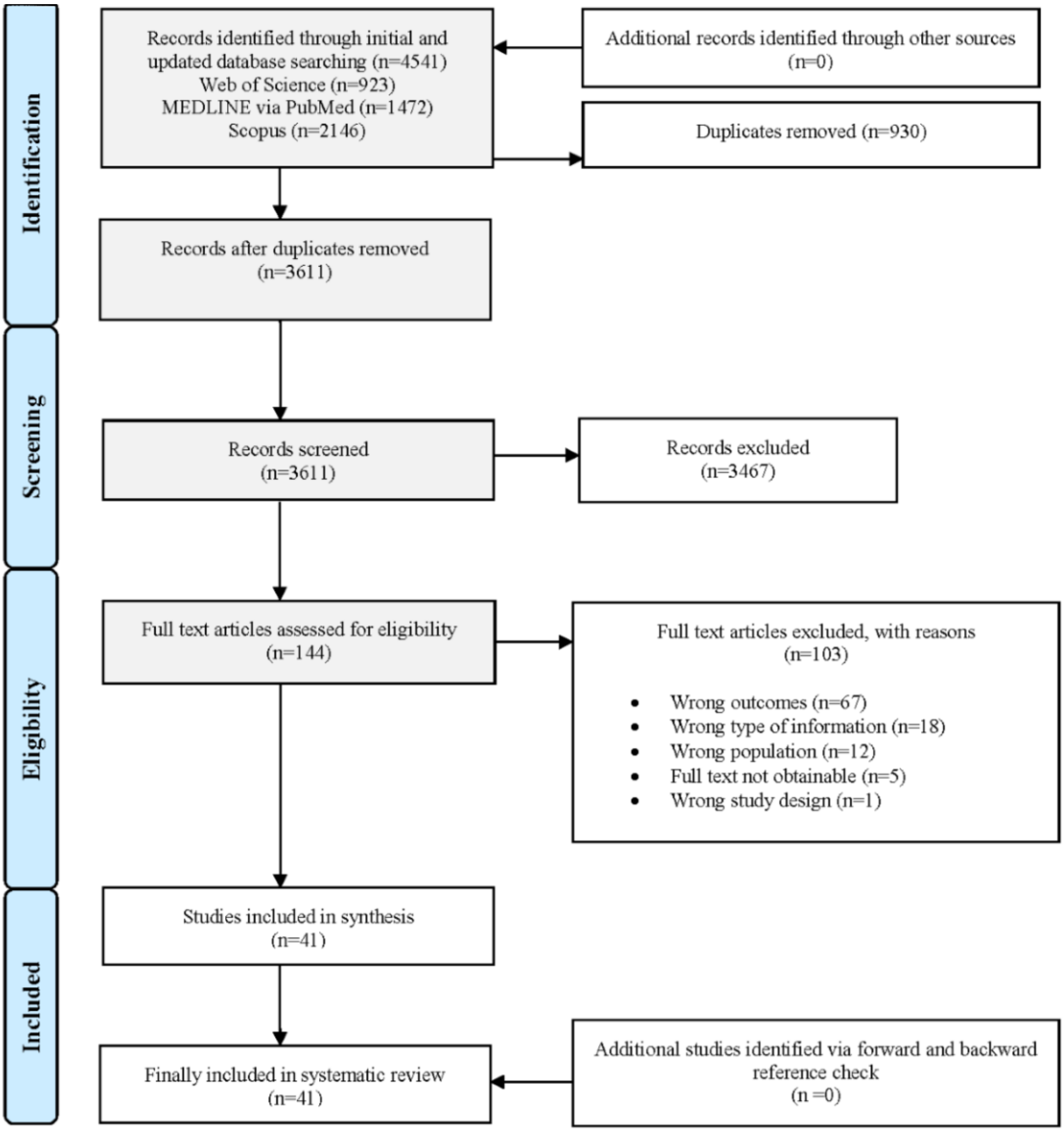
During the collection of the main categories, this study observed that *needs* could be expressed directly and explicitly (eg, *need for reduced information*), or more indirectly (implicitly), by naming distinct barriers (eg, *the overabundance of information*), perceived lacks (eg, *lack of reduced information*), and possible facilitators (eg, *suggesting less information*) during the information search. Where those diverse expressions occurred, they were summarized under the same category to enable the compilation of different aspects of utterance meaning [75]. Furthermore, categories were summarized by acknowledging information search and seeking behavior as a process [9], beginning with a personal need, leading to the access of the information, utilizing it, and the implications of the information used. Inevitably, emerging categories can overlap in certain aspects.

Results

Flow Diagram

After deduplication, the search retrieved 3611 citations. A total of 144 publications were screened in full text, and 103 publications were excluded. We included 41 studies in the synthesis. The study flow is shown in the preferred reporting items for systematic reviews and meta-analyses [76] (PRISMA) diagram (Figure 2).

Figure 2. Flow diagram.



Characteristics of Included Studies

Type of Studies and Data Collection

Out of the 41 included studies, 25 were quantitative studies, 9 were qualitative, and 7 were mixed methods studies. Data were collected through a survey (n=18); interview (n=5); questionnaire or logbook (review of medical notes and consultation records; n=5); collection of clinical questions by interview or observation (n=4); combined survey, interview, or focus groups (n=4); focus groups alone (n=3); or as a prospective study with electronic data collection (n=2). For a detailed overview of qualitative and mixed methods studies as well as the aims of quantitative studies, their characteristics, recruitment settings, and the outcomes formulated, see Tables 1 and 2.

Table 1. Characteristics of qualitative and mixed methods studies.

Author (Reference)	Aims	Data collection	Data analysis	Sample size	Main results and themes	Type of information
Badran et al 2015 [31]	Explore the viewpoint of FPs[a] on advantages and disadvantages of educational email alerts	Interview	Thematic analysis	15 FPs	Advantages or disadvantages of educational email alerts; knowledge, attitude, and behavior toward email	Email or educational email alerts
Barrett et al 2004 [32]	No clear aim stated	Survey and interview	Descriptive design	88 GPs[b] (survey); 15 GPs (interview)	PDA usage by residents; advantages or disadvantages of PDA use	Information regarding the use of PDAs
Boissin 2005 [33]	Consider the use of computers by GPs, analyze the impact of computerization on information-seeking behavior	Interview	N/A[c]	32 GPs	Computerization; opinions about new technologies; the GPs working environment; information behavior	Internet information
Bryant 2004 [34]	Explore factors that motivate GPs to pursue information	Survey interview and focus group	N/A	19 GPs (interview); 39 GPs (focus group)	Information needs and seeking; preferences; attitudes toward libraries; information sources used	Mixed
Cook et al 2013 [35]	Understand barriers and enabling factors influencing physician point-of-care learning and decisions physicians are facing	Focus group	Grounded theory	50 PCPs[d]	Barriers and enabling factors of point-of-care learning	Mixed
Cullen 2002 [36]	Determine the extent of use of the internet for clinical information among FPs	Survey and interview	N/A	294 GPs (survey); 20 GPs (interview)	Use of internet; information sources; type of information sought; search skills; use of critical evaluation of information and impact on decision	Internet information
Ely et al 2000 [37]	Develop a taxonomy of doctors' questions about patient care	Collection of clinical questions (interview)	Observation and text analysis	103 FPs	Classify clinical questions	Clinical questions
Ely et al 2002 [38]	Describe obstacles encountered when attempting to answer doctors' questions with evidence	Collection of clinical questions (observation)	Observation and text analysis	103 FPs	Obstacles encountered trying to obtain evidence-based answers to real clinical questions	Clinical questions
Feightner et al 2001 [39]	Explore FPs' perspectives on how to provide evidence-based preventive clinical practice guidelines to physicians on the internet	Focus groups	Thematic analysis	34 FPs	Preferences for disseminating preventative clinical practice guidelines through the internet	Internet information
González-González et al 2007 [40]	Determine information needs of PCPs and describe their information-seeking patterns	Recording consultations and telephone interview	Classification of questions	112 PCPs; 90 PCPs	Number of questions asked, pursued, and answered; type and topic of questions; time spent pursuing answers; information resources used; perceived barriers to search	Mixed
Heintze et al 2005 [41]	Capture the variety of perceptions and intentions to act and attitudes of GPs regarding their own CME[e] behavior	Interviews	Thematic synthesis	30 GPs	Perceptions of CME programs	Mixed (CME aspect)
Janes et al 2005 [42]	Investigate health professionals' attitudes and perceived barriers to using the internet for ongoing professional learning	Cross-sectional study and questionnaire	Inductive approach	175 GPs (survey); 56 GPs (written response)	Barriers to use the internet	Internet information

Author (Reference)	Aims	Data collection	Data analysis	Sample size	Main results and themes	Type of information
Lottridge et al 2007 [43][f]	Investigate differences that impact physicians' needs of clinical evidence on mobile devices	Interviews	Verbal protocol analysis	47 FPs	Effect of specialty on preferences toward handheld presentation of evidence	Mobile information
Schuers et al 2016 [44]	Explore attitudes and behavior of residents in general medicine and GPs when seeking medical information online	Focus groups	Descriptive analysis	35 GPs	Research topic in general medicine; resource selection; seeking process; research context	Internet information
Vaucher et al 2016 [45]	Assess suggestions of practicing physicians for possible improvements of knowledge transition effectiveness into clinical practice	Interviews; focus groups; online questionnaire	Content analysis; grounded theory	4 GPs (interviews); 25 GPs (focus group); 587 GPs (survey)	Barriers in knowledge transition and suggestions to improve implementation	Mixed
Zack et al 2006 [46]	Better understand GPs' information needs and preferences to provide basis for developing better information resources	Questionnaire; interview; review medical notes	Grounded theory	47 GPs	Reasons for difficulties and coping strategies; information sources used	Mixed

[a] FP: family physician.

[b] GP: general physician.

[c] N/A: not applicable.

[d] PCP: primary care physician or practitioner.

[e] CME: continuous medical education.

[f] In this study, family physicians did not make up at least 50% of the population. However, the studies compared different groups of doctors and were included in the analysis to evaluate the potential differences between family physicians and other specialties. There was heterogeneity in the studies regarding the term family physician, general practitioner, or primary care practitioner. We decided to use the term of the original study rather than trying to find a common definition of family physicians for different studies. Types of information grouping were performed by medium or source. Due to a lack of one common definition for the information medium or sources among studies, we summarized those studies sharing a comparable or similar definition. However, some studies needed to be grouped by the addressed content rather than the analyzed medium itself.

Table 2. Characteristics of quantitative studies.

Author (Reference)	Aims	Data collection	Recruitment and sample size[a]	Main outcomes	Type of information
Andrews et al 2005 [47]	Assess information-seeking behaviors and preferences of clinicians	Cross-sectional survey	Network; PCPs[b]=41	Use of and barriers to information resources	Mixed
Bennett et al 2005 [48]	Assess the way FPs[c] use the internet to look for clinical information and how patterns vary from other specialists	Fax survey	Fax database; FPs=457	Usefulness of internet as information resource; search pattern compared with other specialties	Internet information
Bernard et al 2012 [49]	Describe characteristics of GPs[d] using the internet for clinical information search to identify barriers and facilitators to internet use	Cross-sectional survey	Online questionnaire; FPs=721	Internet use for clinical information; obstacles and facilitators for internet use	Internet information
Bjerre et al 2013 [50]	Assess feasibility of using information generated in the context of what could become a “routine” clinical information source	Collection of clinical questions (secondary analysis)	Adjacent study; PCPs=82	Description of type and frequency questions asked (secondary analysis)	Clinical questions
Butzlaff et al 2002 [51]	Demands of GPs toward CME[e] media, the used CME sources, and their efficacy	Survey	Database; FPs=72	Amount of GPs knowledge acquired after studies. Requirements of GPs toward CME. Sources of GPs’ CME. Efficacy of CME. Significance of the internet.	CME media and internet information
Ciarlo et al 2016 [52]	Identify the specific needs of oncologists and GPs attending cancer patients	Survey	Database; GPs=147	Information sources; questions frequently asked; dealing with uncertainty; satisfaction with information; information seeking and time spent; information needs for specific topics	Mixed
Cogdill et al 2000 [53]	Investigate information needs and information seeking in primary care practices serving as educational sites	Interview with follow-up	University; PCPs=15	Frequencies and categories of questions generated during patient encounter	Clinical questions
Davies 2011 [21][f]	Determine information needs of physicians	Survey	Professional bodies; PCPs=256	Information needs. Frequency of formulated questions. Use of computers. Preference in locating evidence. Barriers in accessing electronic information.	Electronic information
Denny et al 2015 [54]	Investigate the use of e-resources within the GPs’ education and training sector	Survey	Organizations; GPs=119	Personal and professional characteristics associated with use of e-resources. Preferred sources. Frequency of use. Factors relevant in selection and use.	E-resources
Dwairy et al 2011 [55]	Explore optimal foraging theory to understand information-seeking behavior of GPs, measure costs, and benefits of information-seeking decisions	Logbook and questionnaire	Within region; GPs=115	Time spent on addressing search for clinical information; preferred information source; success in search	Mixed
Ebell et al 2011 [56]	Identify clinical questions health care professionals have and explore whether questions could be used to drive needs assessment for clinical education programs	Collection of clinical questions (observation and survey)	Personal contact and academic; PCPs=25	Description of type and frequency questions asked	Clinical questions
Koller et al 2001 [57]	Clarify reasons for not consulting the internet and identify alternative sources of information for problem solving during patient care	Cross-sectional survey	Database; PCPs=1103	Internet access and use; reasons for not using	Internet information
Kortekaas et al 2015 [58]	Determine how often and how GP trainees search for answers to clinical queries encountered in daily clinical practice	Logs	University; GP trainees=76	Number of clinical queries and answers pursued and retrieved; outcome on decision making; resources used	Mixed

Author (Reference)	Aims	Data collection	Recruitment and sample size[a]	Main outcomes	Type of information
Kostagiolas et al 2015 [59]	Explore the information-seeking behavior of GPs and their attitudes toward participatory medicine	Survey	Organization; GP=174	Information needs and sources; obstacles when seeking information; perception of participatory medicine	Mixed
Kosteniuk et al 2013 [60]	Determine information resources of FPs to update general medical knowledge and make clinical decisions	Cross-sectional survey	Database; FPs=331	Information sources used and found to be most accessible and relevant to needs	Mixed
Kritz et al 2013 [61][g]	Provide insight on the professional internet use among different subgroups of physicians	Survey	Mixed promotion; GPs=89	Use of online resources; time spent on searches; rate of success; barriers of finding information; search tools	Internet information
Le et al 2016 [62]	Assess GPs' information-seeking behavior, perceived importance of medical information sources and associations with GPs characteristics	Cross-sectional survey	Organizations; GPs=1580	Frequency of use; perceived importance; associations between GP characteristics and use and importance	Mixed
MacWalter et al 2016 [63]	Determine how GPs use online resources in support of their continuing professional development	Cross-sectional survey	Via email; GPs=383	Internet usage and reasons for use; intergroup comparisons	Internet information
Magin et al 2015 [64]	Establish prevalence and associations of GP trainees in consultation information seeking	Cross-sectional survey	Adjacent study; GP trainees=654	In-consultation information seeking from human or nonhuman source	Mixed
Magin et al 2017 [65]	Establish prevalence and associations of GP trainees generation of learning goals	Cross-sectional survey	Adjacent study; GP trainees=1124	Generation of learning goals	Clinical questions
Magrabi et al 2008 [66]	Determine long-term use of an online evidence system in routine clinical practice	Prospective study (data collection via computer log)	N/A; GPs=59	Usage pattern and user group analysis	Online tool
Ruf et al 2008 [67]	Examine GPs' attitudes toward and use of the internet and online CME	Survey	Adjacent study; random sample GPs=351	Internet use and importance; frequency and effectiveness of CME	Internet information (CME aspect)
Schwartz et al 2003 [68]	Determine if FP faculty answer their questions using online resources and the proportions of answers that influenced patient care	Prospective study (data collection via computer log)	Practice center; GPs=3	Characteristics of questions generated and answered; search time; resources used; barriers to use	Internet information
Vollmar et al 2008 [69]	Gain understanding of PCPs' learning media preferences	Survey with follow-up	Postal; PCPs=57	Resources used for CME and effectiveness; demands toward CME media	CME media
Vollmar et al 2009 [70]	Gain understanding of GPs' preferences for different forms of educational media that will meet CME needs	Survey	Adjacent study; PCPs=264	Internet access and utilization; requirements toward CME media	CME media

[a] Sample size describes family physician staff included. If mixed personnel were surveyed or addressed, nonphysicians were excluded from the description and analysis.

[b] PCP: primary care physician or practitioner.

[c] FP: family physician.

[d] GP: general physician.

[e] CME: continuous medical education.

[f] This survey also included a short literature review but was not excluded from the analysis.

[g] In these studies, family physicians did not make up at least 50% of the population. However, the studies compared different groups of doctors and were included in the analysis to evaluate the potential differences between family physicians and other specialties. There was heterogeneity in the studies regarding the term family physician, general practitioner, or primary care practitioner. We decided to use the term of the original study rather than trying to find a common definition of family physicians for different studies. Types of information grouping were performed by medium or source. Due to a lack of one common definition for the information medium or sources among studies, we summarized those studies sharing a comparable or similar definition. However, some studies needed to be grouped by the addressed content rather than the analyzed medium itself.

Type of Information Addressed

Most studies addressed mixed online information sources (n=15) or internet information sources (n=12). Few studies focused on mobile information sources (n=2). One study focused on online information delivered via email (n=1) or online information delivered via an online tool or app (n=1). Some studies focused on the process of generating clinical questions in practice (n=6) or analyzed CME as an online health information resource (n=4).

Synthesis of Studies

A total of 20 subcategories emerged from the coding of the included studies and were summarized into the following 5 main categories:

1. *Individual needs*: Formed to collect FPs' expressed individual personal needs, barriers, or demands toward online information before initiating a search. This category collected diverse statements from FPs when they directly identified a personal need or more indirectly explained the individual barriers met.
2. *Access needs*: Formed to collect aspects, needs, or barriers expressed by FPs during the access of online information.
3. *Quality needs*: Formed to collect aspects, needs, or barriers expressed by FPs toward the quality of online information after being accessed.
4. *Utilization needs*: Formed to collect aspects, needs, or barriers regarding the subsequent utilization of the retrieved information.
5. *Implementation needs*: Formed to collect aspects, needs, or barriers regarding consequences and effects that emerged due to or after utilization of online information.

Meta-analyses of survey results were not possible due to the heterogeneity of methods used in data collection. See Table 3 for details of the main categories, items, and item descriptions.

Table 3. Main categories, themes, and theme descriptions.

Main category and subcategory	Description	Supporting survey results
Individual needs		
CME[a]	There is a need for continuous medical education in practice. There is a need for being kept up-to-date [33-36,39,42,44,45].	89.7% ranked knowledge update as high level of importance as information need or motive [59]; 5-point Likert scale (1-2=low, 4-5=high importance), N=174; 80.4% of FPs[b] use the internet for work-related continuing professional development [63]; survey, N=383.
Digital skill	There is a lack of internet, computer, or digital skills [31,33,36,39,42,45].	Lack of computer or digital literacy skills ranked low level of importance as an obstacle to obtain information by 70.3% and 65.9% of FPs [59]; 5-point Likert scale (1-2=low, 4-5=high importance), N=174.
Collaboration	Collaborations among colleagues or with other clinical fields and experts are important [33-36,38,41,42,44-46].	Colleagues as the preferred aid in clinical decision making among FPs [21]; survey, N=256. Colleagues used as information source by 62.4% of the FPs in making clinical decisions [60]; survey; N=330. For learning activities, German FPs use quality circles (75.7%) and colleagues (58.5%) as preferred information source with significant correlation between utilization and efficacy [70]; survey, N=264.
EBM[c] skill	There is a lack of methodological and scientific principles regarding the practice of EBM. Methodological and scientific skills regarding EBM are needed [34-36,38,41,44,45].	65.0% of FPs see websites with evidence-based summaries as the leading facilitating factor to use the internet for information seeking in clinical practice [49]; survey, N=721. 45.5% of FPs see difficulties in quality appraisal as a leading barrier to using the internet for CME [67]; survey, N=349.
Prefer analogue	Analogue information may be preferred over electronic resources by FPs [35,39,42,45].	Medical textbooks (66.4%) or books or printed journals (86.3%) used by FPs to make specific clinical decisions regarding patient care [49,60]; surveys, N=721 and 330, respectively.
Access needs		
Time	Time to look up or access information is missing. Information access should be quick [21,31,34,35,38-42,44,45].	Lack of time ranked as a leading important obstacle when seeking information [59]; 65.3% high importance, 5-point Likert scale (1-2=low, 4-5=high importance), N=174. 47.0% of FPs ranked lack of time as second most important barrier when searching the internet for clinical information [49]; survey, N=721. FPs spent least amount of time in complex queries compared with other specialties [61]; survey, N=500. Time to search was ranked as the most frequent barrier to look for information [21]; survey, N=256.
Simple	Online resources should facilitate easy access to information. Navigational aspects are important for access to information [32,36,39,42-46].	FP registrars named ease of navigation as a factor relevant to use of e-resources [54]; mean 4.32, SD 0.61, 5-point Likert scale (1=not important, 5=very important), N=119. 61.3% of FPs ranked navigation difficulties first as physician internet barrier [48]; survey, N=457.
Cost	Access to information is expensive. Access to information should be free [32,41,42,44,45].	Cost was ranked as the second highest obstacle when seeking information [59]; 59.2% high importance, 5-point Likert scale (1-2=low, 4-5=high importance), N=174.
Language	Foreign language can be a barrier in the information-seeking process [44].	Language barrier was ranked third by 34.1% of FPs as an obstacle when seeking information in clinical practice [49]; survey, N=721.
Technical	Hardware, software, or technical issues prevent access to information [32,38,39,42].	Most reported difficulties when using online resources for professional development: 62.7% slow internet connection; 49.9% additional software needed; 46.2% access to website restricted; 42.6% problems logging into online resource; and 37.3% internet connection problems other than speed [63]; survey, N=383.
Quality needs		
Credible	Information and the institution offering it should be credible, transparent, and trustworthy. Information should be independent from pharmaceutical firms or industry [32,34-36,38,39,41,44-46].	Reliability is the second most favored attribute regarding tools for CME [70]; 89.8% very important, 3-point ordinal scale (0=unimportant, 2=very important), survey N=264. Pharmaceutical sales representatives are the least used information source by FPs [60]; 4.2%, survey, N=330.

Main category and subcategory	Description	Supporting survey results
Concise	Overabundance of information can result in an ineffective search of information. Information should be preselected and comprehensive to FPs’ relevant topics [31,34,35,38,39,41,42,44-46]. FPs need short and concise summaries of information [35,38,39,41,43-46].	“Too much information to scan” named as leading barrier (47.7%) to internet use for information seeking and identified “evidence-based summaries” and “selected documents” as leading facilitating factors for information seeking (65.0% and 54.4%, respectively) [49]; survey, N=721. “Content filters” perceived as an important tool for information search, identified by 48.0% of FPs [61]; survey; N=89.
Up-to-date	Information should be recent and up-to-date [35,38,39,42,45,46].	“Creation date listed” was identified as an important factor relevant to GPs’[d] use of e-resources [54]; mean 4.22, SD 0.72, 5-point Likert scale (1=not important, 5=very important) survey, N=119.
Specific	There is a need for specific and in-depth information among FPs that is highly variable and dependent on the situation (eg, rare diseases and pediatric doses) [32,34-36,38,39,41,44,46].	66.7% of FPs search for specific patient information, 44.0% of FPs identified the lack of availability of specific information as a barrier to using the internet [48]; survey, N=457.
Utilization needs		
Usability	FPs identify easy navigation and organized content as important for the daily usability of an electronic resource [32,35,38,39,43,46].	—[e]
Science-practice gap	FPs note an existing gap between scientific literature and the questions arising from daily practice [38,45].	—
Doctor-patient-relationship	FPs see implications for the doctor-patient relationship, when information search is done during the patient encounter [33,35,44].	71.0% of FPs name “disturbance of doctor-patient-communication” as a leading reason for not using the internet [57]; survey, N=1103.
Implication needs		
Relevancy for daily practice	Electronic information should be useful or relevant to daily practice and individual setting. Information should aid or improve the process of clinical decision making [32,34,35,38,41,42,44-46].	“Relevant to practice” is rated as a very important requirement and most favored attribute of educational media use by 93.3% of FPs [70]; 3-point ordinal scale (0=unimportant, 2=very important), survey N=264. 27.0% of FPs name “low relevance for clinical practice” as a barrier to using the internet for information seeking, and nearly half of the FPs see more relevancy for clinical practice as a facilitating factor [49]; survey; N=721.
Patient education	Information should be useful for patient education [32,34-36,39,43,44,46].	93.5% of FPs use the internet for obtaining information to give to a patient [63]; survey, N=383.
Justification of practice	FPs search for information to justify practice or clinical decision [34-36,39,41,44,46].	Improvement of clinical decision making and confirmation of decision are among the most frequently named impacts of information search among FP trainees [58]; 25.8% and 22.7% of clinical queries in daily practice; survey; N=76.

[a] CME: continuous medical education.

[b] FP: family physician.

[c] EBM: evidence-based medicine.

[d] GP: general physician.

[e] No substantiating quantitative results are displayed.

Individual Needs

CME

FPs identified a need for CME in everyday practice [33-36,39,42,44,45] but did not rank the internet as the most preferred source [36,49] for obtaining CME-related information. Although the work-related utilization of the internet for CME is quite high [63], FPs appear to prefer personal medical education such as colleagues and quality circles for updating their knowledge [51,67,70].

Digital Skill

FPs mentioned a lack of digital, computer, or internet skills as a potential barrier in obtaining online health information [31,33,36,39,42,45]. However, the lack of digital or technical skills was not mentioned as a leading obstacle to obtaining online information [48,49,57,59,67]. A cross-sectional survey made more precise distinctions in mentioning digital or technical difficulties when using online resources, displaying the variety of digital or technical barriers that can occur when using new technologies (eg, log-in problems and need for additional software) [63].

Collaboration

FPs expressed the need for collaborations with colleagues or experts throughout different disciplines and institutions (practice, hospital, and universities) when seeking information [33-36,38,41,42,44-46]. Quantitative data from surveys support the utilization of colleagues and experts as an important information resource for FPs [21,51,52,55,57-60,62,69,70]. Colleagues were the resource with the highest success rate when obtaining information among FPs, being more efficient than search engines or websites [55]. Young FP registrars named face-to-face contact with educators or colleagues as the second most preferred resource after using e-resources [54].

Evidence-Based Medicine Skill

FPs realized a lack of various skills and competences relating to methods and principles of practicing evidence-based medicine (EBM; eg, literature search and critical appraisal) [34-36,38,41,44,45]. Surveys mentioned the difficulty in obtaining quality appraisals as a hindrance to their use of the internet for CME [67] and identified websites with evidence-based summaries as facilitators of their use of the internet for information seeking in clinical practice [49].

Prefer Analogue

In a few qualitative studies, analogue sources of information were preferred by some FPs over electronic resources [35,39,42,45]. Quantitative studies show varying and inconclusive results concerning the FPs’ most preferred sources of information [21,47,51,52,55,58-60,62,67,69,70].

Access Needs

Time

Lack of time was frequently referred to as a barrier to accessing information. Quick access to information was demanded by FPs in both qualitative studies and surveys [21,31,34,35,38-42,44,45,49,51,57,59,69,70]. FPs were reported as devoting the least amount of time to complex queries, and they are more likely to perceive a lack of time than other specialists [61]. FPs spent 18 min on average on their searches for clinical information [55]. FPs refer to the lack of time as a leading barrier to obtaining information from the internet [49,57,59]. FPs also ranked the attribute *fast* as a leading criterion for the efficient utilization of information [51,69,70].

Simple

FPs mentioned easy access as an important requirement in the process of seeking and obtaining information. Emphasis was laid on simple technological aspects or technological tools that enhanced information access [32,36,39,42-46]. Surveys supported the fact that complex technological procedures appear to be a hindrance to the access of online information [49,63,67]. User friendliness was mentioned as an important requirement in obtaining electronic information [51,69,70]. Another aspect was the identification of navigation difficulties as a barrier to obtaining information from the internet [48] as well as mentioning the ease of navigation as a factor that was highly relevant to FPs when using e-resources [54].

Cost

FPs named costs as a barrier to accessing information. On the one hand, they expressed the need for free access to information, yet on the other hand, they mentioned costs of hardware and software as a hindrance to obtaining information [32,41,42,44,45]. Surveys supported cost and cost-effectiveness as a factor for FPs when using CME [51,62,67,69,70], although no obvious conclusion was drawn from the importance of this factor as a barrier to accessing information in general [49,59].

Language

Qualitative studies rarely mentioned languages as an obstacle for obtaining information [44]. However, surveys identified foreign languages as a possible barrier in the process of seeking information [49,59]. For German FPs, language is of medium importance when using the internet [67,69,70].

Technical

The technical aspects identified as preventing the access of information or displaying a barrier to the process of seeking information were named in several studies covering a wide variety of technical, hardware and software, or internet-related problems [32,38,39,42]. A quantitative survey among Scottish FPs identified several distinct issues, such as a slow connection or incompatible software, when accessing information for CME [63].

Quality Needs

Credible

FPs’ needs regarding the quality of information, trustworthiness, credibility, and transparency of information and the institution generating this information were frequently named in qualitative studies [34-36,38,39,44-46]. Transparency and credibility were often linked to the need for information to be independent from the pharmaceutical industry [32,38,41,44,45]. Quantitative studies supported the need for trustworthy, credible, and transparent information among FPs [48,51,57,60,63,67,69,70].

Concise

FPs cited an overabundance of information as a barrier to the process of searching for specific or relevant information. The internet and other electronic information resources were perceived as containing an untamed *information jungle*, hindering the effectiveness of researching FP-relevant information [31,34,35,38,39,41,42,44-46]. This result was supported by several surveys that addressed the need for concise information or identified *too much* or *confusing* information as an access barrier [49,51,54,57,59,61,67,69,70]. Another aspect identified by several qualitative studies was the FPs’ need for short and concise summaries of information [35,38,39,41,43-46].

Up-to-Date

Another need identified by FPs was the currency of information available [35,38,39,42,45,46], which was chiefly cited by qualitative analyses rather than quantitative studies [54,67].

Specific

FPs seemed to show differing needs for specific information depending on the particular clinical question at hand and the individual patient situation. These needs could cover anything from detailed pediatric drug dosing to diagnostic criteria for rare diseases, but could not be narrowed down to any specific or homogenous topic [32,34-36,38,39,41,44,46]. Therefore, quantitative data were too heterogeneous to present a distinct pattern of those specific needs emerging from surveys. However, a survey listed the unavailability of specific information as a barrier for FPs [48]. It was not an aim of this review to analyze the distinct medical information FPs were searching for, but some of the included studies identified these topics or developed or used the taxonomy of clinically generated questions by FPs [37,40,50,56,64,65,68]. This supported the highly heterogeneous field of clinical questions that could arise from the FPs' daily routine. Cook et al [35] noted that the complexity of clinical questions was an important aspect to consider among FPs.

Utilization Needs

Usability

The most prominent aspect retrieved was the need for easy navigation and an organized display of structured content [32,35,38,39,43,46]. Minor aspects retrieved from some studies also mentioned the need for short and summarized information [35,39] and aspects regarding mobile or tablet resources such as physical size, screen requirements, or applications used [32,43,45]. Quantitative evidence identifying needs as suitable for daily practice utilization was sparse [70].

Science-Practice Gap

Few qualitative studies mentioned that scientific literature failed to address and reflect on the relevant problems emerging from daily practice, omitting the connection of academic centers to daily practice [38,45]. None of the included surveys directly measured this aspect. Few surveys report that FPs perceived a lack of specific information [48] or the low relevance to clinical practice [49] as a barrier to searching for information on the internet.

Doctor-Patient Relationship

Qualitative studies suggested considering the setting of the FP encounter with the patient, and possible positive and negative implications on the doctor-patient relationship, as a consideration when information searches were conducted during the encounter [33,35,44]. One focus group study, in particular, named the complexity of questions that arose in general practice as a barrier to searching for information [35]. An older survey of Swiss doctors identified the interruption of doctor-patient communication as a reason for not using the internet [57].

Implication Needs

Relevance for Daily Practice

One important implication for the FPs' everyday practice was a reported lack of usefulness and relevance of electronic resources in the daily clinical routine. FPs noted that information should be applicable to their specific daily situations, rather than general guidelines and recommendations [32,34,35,38,41,42,44-46]. Surveys supported the need for information relevance to daily practice [51,67,69,70] or identified low relevance as a barrier to information seeking [49]. Furthermore, surveys reported the need for information to make improved clinical decisions [58,63].

Patient Education

An important viewpoint of many FPs was the usefulness of retrieved information for patient education or the need for information supporting the patients' involvement in the process of explanation, identified by a number of qualitative studies [32,34-36,39,43,44,46]. A survey of Scottish FPs reported that over 90% of them used the internet to obtain information for the patient [63] or advise patients on internet health resources [47].

Justification of Practice

Qualitative studies also showed that general physicians (GPs) searched and used information in everyday practice to reaffirm preexisting knowledge or to justify their clinical decisions [34-36,39,41,44,46]. The search strategies of FP trainees also demonstrated the impact on clinical decision making or the confirmation of a diagnosis [58]. Surveys that developed or relied on the taxonomy arising from FPs' daily practice also supported this finding. The most common question types could be classified according to the categories *diagnosis* and *treatment* [37,40,50,56,64,65,68]. Surveys also show that the topics *diagnosis* and *treatment* were important information-seeking motives or information needs among FPs [52,53,59].

Discussion

Principal Findings

This study presented 5 main aspects of FPs' needs toward online health information:

1. Several individual needs exist for FPs before online information is accessed, such as the need for digital and EBM skills, preference for analogue information or a desire for CME, and the need for interspecialist collaborations.
2. Needs that are connected with the access of online information, such as simple access, technical barriers, a good cost-benefit ratio, or suitable languages.
3. Needs that address aspects of quality itself, for instance, credible and recent information. The most interesting aspects of quality revealed a converse need for concise information, on the one hand, as well as the need for specific in-depth information, on the other hand.
4. Needs that are concerned with the feasible utilization of obtained information, such as the suitability of information to the distinct and unique situations in family practice.
5. Needs that reflect the subsequent implications of using information that is tailored to FP practice, clinical judgment and decision making, and patient education as well as providing additional value to the FPs' future practice.

Comparison With Prior Work

The impact of the internet on the information needs of primary care was reviewed in 1999 and identified FPs' need to manage information overload as well as the need for specific and simple information [5]. This study confirms these findings within the quality needs category and thus confirms prior work as still valid. Rural health professionals have information needs directly relating to patient care and therapy, and they cited a lack of time or technological literacy as barriers to obtaining information [6]. Our study supports the relevance of patient-related needs mainly in the utilization and implication needs category. It also confirms time and technological aspects in the access needs category.

Dawes and Sampson [7] noted heterogeneous information-seeking behavior among doctors in 2003 and asked for careful planning in delivering useful, relevant, and fast information to physicians, supported by our findings within the main categories of utilization and implication needs. A noteworthy review from 2006 identified the information-seeking obstacles to primary care physicians in the context of established EBM processes [8]: (1) acknowledging an information gap, (2) formulating a question, (3) seeking relevant information, (4) formulating an answer, and (5) applying the answer to patient care. Although our review is able to confirm most of the barriers reported by Coumo and Meijman [8], it intends to present an adapted classification of the steps necessary for FPs to obtain such information.

The FPs' information needs cannot be completely met by only providing high-quality information through newly tailored online sources. New content, new technologies, or new systems must address seeking competencies, strategies of utilization, and the implications generated in family practice, as our results revealed. The internet's role in needs, the information-seeking patterns, and the sources utilized was partly reviewed by Davies [9]. In agreement with some of our subcategories, prominent barriers identified in information searches were lack of time, lack of information technology skills, and lack of search skills, although needs were often related to diagnosis and therapy. Physicians' information needs are often related to diagnosis, therapy, and patient care, as confirmed by literature reviews from 2010 and 2013 [10,11]. The most recent systematic review in this field of work was conducted by Del Fiol et al [12] in 2014, confirming that clinicians raise questions about patient care. Although our study confirms the findings of these more recent reviews, none of them have been exclusively focused on FPs. Therefore, a substantiating comparison remains complex.

What Is New and Where to Go From Here?

Despite confirming prior work, this study seeks to highlight possible future work emerging from the results presented. The main categories and subcategories indicate that needs toward online information by FPs seem to be closely associated with CME and EBM. Evidence suggests that EBM interventions improve short-term knowledge, but there is little evidence of a change in long-term knowledge, attitudes, or clinical practice [77]. Despite technological advancements, half of the clinical questions still seem to be unanswered at the point of care [12]. No study has directly measured the effects of these interventions on patients' outcomes or FPs' behavior [8]. The inability to search for the literature and critically appraise the content—both inevitable steps of EBM [78]—were identified as barriers to obtaining information in the first place by this study. We suggest that future work should focus on these intermingling aspects of information need, CME, EBM, and daily routine in the primary care setting. It should not abandon the implications and effects on FPs' behavior or patient outcomes that occur after an information search or when a question is not pursued.

Searching for and critically appraising primary literature in a short amount of time remains a major obstacle in primary care, urging FPs to express needs for concise secondary, credible, free, and simple information that also provides valuable and specific medical information. This converse need for short and concise, but also in-depth, information for an FP, in our opinion, has neither been met by new online information platforms nor by science contributing to the information translation with relevant research into the FPs' daily practice. An FP's need for information rarely starts with a scientific definition of an illness or an update on epidemiology, but with a specific question on individual patients and with direct impact on the situation presented during the consultation. The vast amount of information available across multiple platforms and sources emerges as an obstacle to both initiating and pursuing a clinical question in the FPs' daily practice and consultation. Lack of time remains a major obstacle to information retrieval among FPs, despite the abundance of online information. This emphasizes the fact that online information has not yet fully evolved to satisfy the needs of FPs, explaining that FPs may still prefer colleagues and analogue information in many situations over digital solutions, as it is free, delivered by specialist colleagues, fast, simple, and concise.

Methodological Strengths and Limitations

To our knowledge, this is the first systematic review that analyses the available qualitative and quantitative evidence focused solely on FPs using online health information. As our search was not limited to a specific study design, we feel it is unlikely, but not impossible, that further relevant publications are available. However, the heterogeneity among countries and their unique health care systems made it challenging to find a common term for *family physician*, *family practitioner*, *general practitioner*, and *primary care physician* among the studies included. Both the differences in health care and educational systems can result in a heterogeneous study population of *family physicians*. As this study excluded works from countries with a completely different primary care or health care system or far less developed technological infrastructure than the majority of those in the included studies, this review may display a bias in this aspect of selection.

There is no established methodological approach for synthesizing both qualitative and quantitative data [79], and a variety of methods seems plausible [74]. We, therefore, used specific steps for quality appraisal and synthesis of the studies by following the thematic synthesis by Thomas and Harden [71], referring to the study by Möhler and Meyer [72], and applying the well-established (eg, [28-30]) critical appraisal tools CASP [26] and AXIS [27]. Despite independent review from 2 scientists, the critical appraisal and the reported items cannot cover all aspects of the heterogeneous body of evidence. We neither wanted to unduly appraise nor indecently criticize the studies' quality or the authors' contribution to the scientific community. The final critical appraisal must remain with the scientist using the included original study. According to Hong et al [74], when addressing one overall review question, as is the case with this review, a sequential study synthesis design is applicable. Despite presenting qualitative and quantitative results parallel to our results, we first synthesized qualitative themes and then collected evidence from quantitative studies to support and enrich these developed themes.

We tried to minimize an aspect we named *technological bias* by limiting studies from the years spanning 2000 to 2020. Through the chosen search terms as well as the established exclusion criteria, we sought to ensure that only studies regarding electronic information were included, when the internet and computers were broadly available in most countries. Still, the technical developments of the last 20 years have been expeditious and have resulted in a rapidly developing infrastructure, hardware, and software environment. We noticed the resulting variety of electronic information, ranging from CD-ROM to very recent online databases. Therefore, a small technological bias remains, especially due to older studies that analyzed technological information systems and corresponding seeking behaviors, which are generally no longer used or even obsolete in 2020 (eg, CD-ROM and Palm OS).

Conclusions: FPs’ View

Although technology and infrastructure, methods, and sources of information retrieval have changed, the needs and barriers of FPs to information seeking and retrieval have not. The question arises, why do technological advancements not succeed in fulfilling the information needs of FPs?

We propose the following two main answers to this question:

1. Human sources of information, such as colleagues, play an important role. FPs are the center of an afferent information flow, as they receive health information from hospitals and other specialists. The FP provides primary care for patients presenting with a variety of illnesses and questions. There is a tension field for the FPs as information givers with an efferent information flow toward the patient. FPs need to develop coping strategies to tackle the demands met in this center of bidirectional information flow by seeking CME and EBM, both instruments to improve knowledge and information retrieval.
2. FPs acknowledge their need for digital skills to search and find the information needed in the online information jungle. It is interesting to note that the methods used for providing this information have come to signify the transition from the analogue to the digital era, although the way of presenting this information for the FPs’ daily work has not yet kept up with this transition.

This review aims to contribute to a (1) FP-specific and (2) an updated systematic body of research that also sought to analyze (3) the influencing factors affecting needs and requirements for online information in primary care. This study concludes that FPs show specific needs for online information due to their daily routine and broad working environment. Future information resources, whether online or analogue, must address the needs emerging from the primary care setting as well as rethink the way in which information is adapted to the needs of the digital age. This requires not only the development and implementation of new information systems but also the evaluation of their effects on both physicians and patients. Finally, science should also rethink the way online medical information is disseminated, adapted, and translated into daily practice.

Acknowledgments

The authors thank Kathrin Schnitzius for helping to prepare the manuscript and for revising the referenced literature. The authors thank Andrea C Schöpf for the initial help with finding a fitting methodological approach and the critical appraisal items. The authors thank Angelina Kloz for critically revising the main categories, subsequent items, and item descriptions. The authors want to thank all authors of the included original studies for their work that made this review possible. This review is part of the General practitioner And Patient communication trial (DRKS00014279, German Clinical Trial Register). The trial is funded by Innovationsausschuss of the Gemeinsamer Bundesausschuss Grant no: 01NVF17010. The funding agency has no role in any aspect of the trial, be this in planning or conducting the trial or in the analysis, interpretation, reporting, or dissemination of study results.

Authors' Contributions

PK was responsible for developing the search terms and conducting the database search. PK served as a screener for the title or abstract and full texts, developed the methodological approach for the search and the critical appraisal, conducted the critical appraisal, synthesized the studies’ results, and prepared and drafted the manuscript. LF was responsible for developing the search terms and conducting the database search. LF served as a screener for the title or abstract and full texts, developed the methodological approach for the search and the critical appraisal, conducted the critical appraisal, and critically revised the manuscript first draft. JT critically revised the manuscript and synthesized the critical appraisal results. KW developed the methodological approach, critically revised the manuscript, and helped to develop tables of this paper. RM critically revised the manuscript and supervised the methodological approach. MS critically revised the manuscript and edited the manuscript as a native speaker. AM and SVR critically revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and search characteristics.
[DOCX File , 19 KB - jmir_v22i12e18816_app1.docx]

Multimedia Appendix 2
Methodological quality of qualitative and mixed methods studies.
[DOC File , 94 KB - jmir_v22i12e18816_app2.doc]

Multimedia Appendix 3
Methodological quality of quantitative studies.
[DOC File , 149 KB - jmir_v22i12e18816_app3.doc]

References

1. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? Ann Intern Med 1985 Oct;103(4):596-599. [doi: 10.7326/0003-4819-103-4-596] [Medline: 4037559]
2. Gorman PN. Information needs of physicians. J Am Soc Inf Sci 1995 Dec;46(10):729-736. [doi: 10.1002/(sici)1097-4571(199512)46:10<729::aid-asi3>3.0.co;2-2]
3. Verhoeven A, Boerma E, Meyboom-de BJ. Use of information sources by family physicians: a literature survey. Bull Med Libr Assoc 1995 Jan;83(1):85-90 [FREE Full text] [Medline: 7703946]
4. Smith R. What clinical information do doctors need? Br Med J 1996 Oct 26;313(7064):1062-1068 [FREE Full text] [doi: 10.1136/bmj.313.7064.1062] [Medline: 8898602]
5. Westberg EE, Miller RA. The basis for using the Internet to support the information needs of primary care. J Am Med Inform Assoc 1999;6(1):6-25 [FREE Full text] [doi: 10.1136/jamia.1999.0060006] [Medline: 9925225]
6. Dorsch J. Information needs of rural health professionals: a review of the literature. Bull Med Libr Assoc 2000 Oct;88(4):346-354 [FREE Full text] [Medline: 11055302]
7. Dawes M, Sampson U. Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians. Int J Med Inform 2003 Aug;71(1):9-15. [doi: 10.1016/s1386-5056(03)00023-6] [Medline: 12909153]
8. Coumou H, Meijman F. How do primary care physicians seek answers to clinical questions? A literature review. J Med Libr Assoc 2006 Jan;94(1):55-60 [FREE Full text] [Medline: 16404470]
9. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J 2007 Jun;24(2):78-94 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00713.x] [Medline: 17584211]
10. Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. Health Info Libr J 2010 Mar;27(1):2-10 [FREE Full text] [doi: 10.1111/j.1471-1842.2010.00883.x] [Medline: 20402799]
11. Clarke MA, Belden JL, Koopman RJ, Steege LM, Moore JL, Canfield SM, et al. Information needs and information-seeking behaviour analysis of primary care physicians and nurses: a literature review. Health Info Libr J 2013 Sep;30(3):178-190 [FREE Full text] [doi: 10.1111/hir.12036] [Medline: 23981019]
12. Del Fiol G, Workman TE, Gorman PN. Clinical questions raised by clinicians at the point of care: a systematic review. JAMA Intern Med 2014 May;174(5):710-718. [doi: 10.1001/jamainternmed.2014.368] [Medline: 24663331]
13. Masters K. For what purpose and reasons do doctors use the Internet: a systematic review. Int J Med Inform 2008 Jan;77(1):4-16. [doi: 10.1016/j.ijmedinf.2006.10.002] [Medline: 17137833]
14. Wald HS, Dube CE, Anthony DC. Untangling the Web—the impact of Internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]
15. Marstedt G. Das Internet: Auch Ihr Ratgeber für Gesundheitsfragen? Bevölkerungsumfrage zur Suche von Gesundheitsinformationen im Internet und zur Reaktion der Ärzte. bertelsmann-stiftung.de. Gütersloh, Germany: Bertelsmann Stiftung; 2017. URL: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/VV_Studie_Das-Internet-auch-Ihr-Ratgeber_Befragung.pdf [accessed 2020-12-14]
16. Dutton W, Blank G, Groselj D. Cultures of the Internet: The Internet in Britain Oxford Internet Survey 2013 Report. Oxford Internet Surveys. Oxford, United Kingdom: University of Oxford; 2013. URL: http://oxis.oii.ox.ac.uk/wp-content/uploads/2014/11/OxIS-2013.pdf [accessed 2020-12-14]
17. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
18. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb 23;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
19. Metzger MJ, Rieh SY, Hilligoss B, Flanagin AJ, Eastin MS. Credibility of health information and digital media: new perspectives and implications for youth. In: Digital Media,Youth, and Credibility. Cambridge, London: MIT Press; 2008:123-154.

20. Braun L, Wiesman F, van den Herik HJ, Hasman A. Avoiding literature overload in the medical domain. Stud Health Technol Inform 2006;124:497-502. [Medline: 17108567]
21. Davies K. Information needs and barriers to accessing electronic information: hospital-based physicians compared to primary care physicians. J Hosp Librariansh 2011 Jul;11(3):249-260. [doi: 10.1080/15323269.2011.587103]
22. Individuals using the Internet (% of population). World Bank Group. 2020. URL: https://data.worldbank.org/indicator/IT.NET.USER.ZS [accessed 2020-12-14]
23. Better Systematic Review Management. Covidence. URL: https://www.covidence.org/home [accessed 2020-12-14]
24. Schmidt B, Durao S, Toews I, Bavuma C, Hohlfeld A, Nury E, et al. Screening strategies for hypertension. Cochrane Database Syst Rev 2020 May 7;5:CD013212. [doi: 10.1002/14651858.CD013212.pub2] [Medline: 32378196]
25. Williams A, Bird M, Hardcastle S, Kirschbaum M, Ogden K, Walters J. Exercise for reducing falls in people living with and beyond cancer. Cochrane Database Syst Rev 2018 Oct 15;10:CD011687 [FREE Full text] [doi: 10.1002/14651858.CD011687.pub2] [Medline: 30320433]
26. Casp Checklist: 10 Questions to Help You Make Sense of a Systematic Review. Critical Appraisal Skills Programme. 2018. URL: https://casp-uk.net/wp-content/uploads/2018/03/CASP-Systematic-Review-Checklist-2018_fillable-form.pdf [accessed 2020-06-30]
27. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 2016 Dec 8;6(12):e011458 [FREE Full text] [doi: 10.1136/bmjopen-2016-011458] [Medline: 27932337]
28. Guyatt G, Sackett D, Cook D. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. what were the results and will they help me in caring for my patients? Evidence-based medicine working group. J Am Med Assoc 1994 Jan 5;271(1):59-63. [doi: 10.1001/jama.271.1.59] [Medline: 8258890]
29. Wakida EK, Talib ZM, Akena D, Okello ES, Kinengyere A, Mindra A, et al. Barriers and facilitators to the integration of mental health services into primary health care: a systematic review. Syst Rev 2018 Nov 28;7(1):211 [FREE Full text] [doi: 10.1186/s13643-018-0882-7] [Medline: 30486900]
30. de Cock C, van Velthoven M, Milne-Ives M, Mooney M, Meinert E. Use of apps to promote childhood vaccination: systematic review. JMIR Mhealth Uhealth 2020 May 18;8(5):e17371 [FREE Full text] [doi: 10.2196/17371] [Medline: 32421684]
31. Badran H, Pluye P, Grad R. Advantages and disadvantages of educational email alerts for family physicians: viewpoint. J Med Internet Res 2015 Feb 27;17(2):e49 [FREE Full text] [doi: 10.2196/jmir.3773] [Medline: 25803184]
32. Barrett JR, Strayer SM, Schubart JR. Assessing medical residents' usage and perceived needs for personal digital assistants. Int J Med Inform 2004 Feb;73(1):25-34. [doi: 10.1016/j.ijmedinf.2003.12.005] [Medline: 15036076]
33. Boissin F. Information-seeking behaviour and use of the Internet by French general practitioners: a qualitative study. Health Info Libr J 2005 Sep;22(3):173-181 [FREE Full text] [doi: 10.1111/j.1471-1842.2005.00573.x] [Medline: 16109145]
34. Bryant S. The information needs and information seeking behaviour of family doctors. Health Info Libr J 2004 Jun;21(2):84-93 [FREE Full text] [doi: 10.1111/j.1471-1842.2004.00490.x] [Medline: 15191599]
35. Cook DA, Sorensen KJ, Wilkinson JM, Berger RA. Barriers and decisions when answering clinical questions at the point of care: a grounded theory study. JAMA Intern Med 2013 Nov 25;173(21):1962-1969. [doi: 10.1001/jamainternmed.2013.10103] [Medline: 23979118]
36. Cullen R. In search of evidence: family practitioners' use of the Internet for clinical information. J Med Libr Assoc 2002 Oct;90(4):370-379 [FREE Full text] [Medline: 12398243]
37. Ely J, Osheroff J, Gorman P, Ebell M, Chambliss M, Pifer E, et al. A taxonomy of generic clinical questions: classification study. Br Med J 2000 Aug 12;321(7258):429-432 [FREE Full text] [doi: 10.1136/bmj.321.7258.429] [Medline: 10938054]
38. Ely J, Osheroff J, Ebell M, Chambliss M, Vinson D, Stevermer J, et al. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. Br Med J 2002 Mar 23;324(7339):710 [FREE Full text] [doi: 10.1136/bmj.324.7339.710] [Medline: 11909789]
39. Feightner J, Marshall J, Sangster L, Wathen C, Quintana Y. Evidence-based preventive practice guidelines. Qualitative study of useful resources on the Internet. Can Fam Physician 2001 Aug;47:1577-1583 [FREE Full text] [Medline: 11561334]
40. González-González AI, Dawes M, Sánchez-Mateos J, Riesgo-Fuertes R, Escortell-Mayor E, Sanz-Cuesta T, et al. Information needs and information-seeking behavior of primary care physicians. Ann Fam Med 2007;5(4):345-352 [FREE Full text] [doi: 10.1370/afm.681] [Medline: 17664501]
41. Heintze C, Matysiak-Klose D, Braun V. Perception of continuous medical education by primary care physicians. Z Arztl Fortbild Qualitatssich 2005;99(7):437-442. [Medline: 16277059]
42. Janes R, Arroll B, Buetow S, Coster G, McCormick R, Hague I. Rural New Zealand health professionals' perceived barriers to greater use of the internet for learning. Rural Remote Health 2005;5(4):436 [FREE Full text] [Medline: 16336054]
43. Lottridge DM, Chignell M, Danicic-Mizdrak R, Pavlovic NJ, Kushniruk A, Straus SE. Group differences in physician responses to handheld presentation of clinical evidence: a verbal protocol analysis. BMC Med Inform Decis Mak 2007 Jul 26;7:22 [FREE Full text] [doi: 10.1186/1472-6947-7-22] [Medline: 17655759]

44. Schuers M, Griffon N, Kerdelhue G, Foubert Q, Mercier A, Darmoni SJ. Behavior and attitudes of residents and general practitioners in searching for health information: from intention to practice. Int J Med Inform 2016 May;89:9-14. [doi: 10.1016/j.ijmedinf.2016.02.003] [Medline: 26980354]
45. Vaucher C, Bovet E, Bengough T, Pidoux V, Grossen M, Panese F, et al. Meeting physicians' needs: a bottom-up approach for improving the implementation of medical knowledge into practice. Health Res Policy Syst 2016 Jul 18;14(1):49 [FREE Full text] [doi: 10.1186/s12961-016-0120-5] [Medline: 27431911]
46. Zack P, Devile C, Clark C, Surtees R. Understanding the information needs of general practitioners managing a rare genetic disorder (osteogenesis imperfecta). Community Genet 2006;9(4):260-267. [doi: 10.1159/000094475] [Medline: 17003536]
47. Andrews J, Pearce K, Ireson C, Love M. Information-seeking behaviors of practitioners in a primary care practice-based research network (PBRN). J Med Libr Assoc 2005 Apr;93(2):206-212 [FREE Full text] [Medline: 15858623]
48. Bennett NL, Casebeer LL, Kristofco R, Collins BC. Family physicians' information seeking behaviors: a survey comparison with other specialties. BMC Med Inform Decis Mak 2005 Mar 22;5:9 [FREE Full text] [doi: 10.1186/1472-6947-5-9] [Medline: 15784135]
49. Bernard E, Arnould M, Saint-Lary O, Duhot D, Hebbrecht G. Internet use for information seeking in clinical practice: a cross-sectional survey among French general practitioners. Int J Med Inform 2012 Jul;81(7):493-499. [doi: 10.1016/j.ijmedinf.2012.02.001] [Medline: 22425281]
50. Bjerre LM, Paterson NR, McGowan J, Hogg W, Campbell CM, Viner G, et al. What do primary care practitioners want to know? A content analysis of questions asked at the point of care. J Contin Educ Health Prof 2013;33(4):224-234. [doi: 10.1002/chp.21191] [Medline: 24347101]
51. Butzlaff M, Koneczny N, Floer B, Vollmar H, Lange S, Kunstmann W, et al. Family physicians, the internet and new knowledge. Utilization and judgment of efficiency of continuing education media by general physicians and internists in family practice. Med Klin (Munich) 2002 Jul 15;97(7):383-388. [doi: 10.1007/s00063-002-1170-8] [Medline: 12168475]
52. Ciarlo G, Liebl P, Zell J, Fessler J, Koester M, Ruetters D, et al. Information needs of oncologists, general practitioners and other professionals caring for patients with cancer. Eur J Cancer Care (Engl) 2016 Nov;25(6):1015-1023. [doi: 10.1111/ecc.12557] [Medline: 27550233]
53. Cogdill KW, Friedman CP, Jenkins CG, Mays BE, Sharp MC. Information needs and information seeking in community medical education. Acad Med 2000 May;75(5):484-486. [doi: 10.1097/00001888-200005000-00020] [Medline: 10824774]
54. Denny B, Chester A, Butler M, Brown J. Australian GP registrars' use of e-resources. Educ Prim Care 2015 Mar;26(2):79-86. [doi: 10.1080/14739879.2015.11494317] [Medline: 25898296]
55. Dwairy M, Dowell AC, Stahl J. The application of foraging theory to the information searching behaviour of general practitioners. BMC Fam Pract 2011 Aug 23;12:90 [FREE Full text] [doi: 10.1186/1471-2296-12-90] [Medline: 21861880]
56. Ebell MH, Cervero R, Joaquin E. Questions asked by physicians as the basis for continuing education needs assessment. J Contin Educ Health Prof 2011;31(1):3-14. [doi: 10.1002/chp.20095] [Medline: 21425354]
57. Koller M, Grütter R, Peltenburg M, Fischer J, Steurer J. Use of the Internet by medical doctors in Switzerland. Swiss Med Wkly 2001 May 5;131(17-18):251-254. [Medline: 11420822]
58. Kortekaas MF, Bartelink ME, Boelman L, Hoes AW, de Wit NJ. General practice trainees' information searching strategies for clinical queries encountered in daily practice. Fam Pract 2015 Oct;32(5):533-537. [doi: 10.1093/fampra/cmv046] [Medline: 26089299]
59. Kostagiolas P, Martzoukou K, Kakavitsas F, Niakas D. The participatory medicine attitudes of general practitioners in greece: an information behaviour perspective. In: Kurbanoglu S, Boustany J, Spiranec S, editors. Information Literacy: Moving Toward Sustainability. Switzerland: Springer International Publishing; 2015:383-393.
60. Kosteniuk JG, Morgan DG, D'Arcy CK. Use and perceptions of information among family physicians: sources considered accessible, relevant, and reliable. J Med Libr Assoc 2013 Jan;101(1):32-37 [FREE Full text] [doi: 10.3163/1536-5050.101.1.006] [Medline: 23405045]
61. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. J Med Internet Res 2013 Jun 26;15(6):e122 [FREE Full text] [doi: 10.2196/jmir.2436] [Medline: 23803299]
62. Le JV, Pedersen LB, Riisgaard H, Lykkegaard J, Nexøe J, Lemmergaard J, et al. Variation in general practitioners' information-seeking behaviour - a cross-sectional study on the influence of gender, age and practice form. Scand J Prim Health Care 2016 Dec;34(4):327-335 [FREE Full text] [doi: 10.1080/02813432.2016.1249057] [Medline: 27804315]
63. MacWalter G, McKay J, Bowie P. Utilisation of internet resources for continuing professional development: a cross-sectional survey of general practitioners in Scotland. BMC Med Educ 2016 Jan 21;16:24 [FREE Full text] [doi: 10.1186/s12909-016-0540-5] [Medline: 26791566]
64. Magin P, Morgan S, Wearne S, Tapley A, Henderson K, Oldmeadow C, et al. GP trainees' in-consultation information-seeking: associations with human, paper and electronic sources. Fam Pract 2015 Oct;32(5):525-532. [doi: 10.1093/fampra/cmv047] [Medline: 26089297]
65. Magin P, Tapley A, Davey A, Morgan S, Holliday E, Ball J, et al. General practitioner trainees' in-consultation generation of clinical questions for later answering: prevalence and associations. Fam Pract 2017 Sep 1;34(5):599-605. [doi: 10.1093/fampra/cmx021] [Medline: 28369454]

66. Magrabi F, Westbrook JI, Kidd MR, Day RO, Coiera E. Long-term patterns of online evidence retrieval use in general practice: a 12-month study. J Med Internet Res 2008 Mar 19;10(1):e6 [FREE Full text] [doi: 10.2196/jmir.974] [Medline: 18353750]
67. Ruf D, Berner M, Kriston L, Maier I, Härter M. General practitioners online: the conditions are good, but use of the Internet for continuing medical education found to be poor. Z Evid Fortbild Qual Gesundhwes 2008;102(5):291-297. [doi: 10.1016/j.zefq.2008.02.023] [Medline: 19006916]
68. Schwartz K, Northrup J, Israel N, Crowell K, Lauder N, Neale A. Use of on-line evidence-based resources at the point of care. Fam Med 2003 Apr;35(4):251-256. [Medline: 12729308]
69. Vollmar H, Ostermann T, Hinz A, Rieger M, Butzlaff M. Primary care physicians, internet and educational media. Preferences, usages and appraisal in a 6-year comparison. Med Klin (Munich) 2008 Jun 15;103(6):425-432. [doi: 10.1007/s00063-008-1055-6] [Medline: 18548212]
70. Vollmar HC, Rieger MA, Butzlaff ME, Ostermann T. General Practitioners' preferences and use of educational media: a German perspective. BMC Health Serv Res 2009 Feb 16;9:31 [FREE Full text] [doi: 10.1186/1472-6963-9-31] [Medline: 19220905]
71. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
72. Möhler R, Meyer G. Attitudes of nurses towards the use of physical restraints in geriatric care: a systematic review of qualitative and quantitative studies. Int J Nurs Stud 2014 Feb;51(2):274-288. [doi: 10.1016/j.ijnurstu.2013.10.004] [Medline: 24176718]
73. Mayring P. Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution. Social Science Open Access Repository. 2014. URL: http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173 [accessed 2020-12-14]
74. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev 2017 Mar 23;6(1):61 [FREE Full text] [doi: 10.1186/s13643-017-0454-2] [Medline: 28335799]
75. Carston R. The explicit/implicit distinction in pragmatics and the limits of explicit communication. Int Rev Pragmat 2009;1(1):35-62. [doi: 10.1163/187731009x455839]
76. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
77. Simons MR, Zurynski Y, Cullis J, Morgan MK, Davidson AS. Does evidence-based medicine training improve doctors' knowledge, practice and patient outcomes? A systematic review of the evidence. Med Teach 2019 May;41(5):532-538. [doi: 10.1080/0142159X.2018.1503646] [Medline: 30328793]
78. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. Br Med J 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
79. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy 2005 Jan;10(1):45-53. [doi: 10.1177/135581960501000110] [Medline: 15667704]

Abbreviations

AXIS: appraisal tool for cross-sectional studies
CASP: critical appraisal skills program
CME: continuous medical education
EBM: evidence-based medicine
FP: family physician
GP: general physician
MeSH: Medical Subject Headings

Edited by G Eysenbach; submitted 20.03.20; peer-reviewed by R Koch, C SteeleGray, J Li; comments to author 22.06.20; revised version received 20.07.20; accepted 11.11.20; published 30.12.20.

Please cite as:
van der Keylen P, Tomandl J, Wollmann K, Möhler R, Sofroniou M, Maun A, Voigt-Radloff S, Frank L
The Online Health Information Needs of Family Physicians: Systematic Review of Qualitative and Quantitative Studies
J Med Internet Res 2020;22(12):e18816
URL: http://www.jmir.org/2020/12/e18816/
doi:10.2196/18816
PMID:33377874

Viewpoint

COVID-19 Contact-Tracing Apps: Analysis of the Readability of Privacy Policies

Melvyn Zhang[1], MBBS, MRCPsych; Aloysius Chow[1], BPsy (Hons); Helen Smith[1], DM, FFPHM

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

Corresponding Author:
Melvyn Zhang, MBBS, MRCPsych
Family Medicine and Primary Care
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road Level 18
Clinical Sciences Building
Singapore, 308322
Singapore
Phone: 65 63892504
Email: melvynzhangweibin@gmail.com

Abstract

Apps that enable contact-tracing are instrumental in mitigating the transmission of COVID-19, but there have been concerns among users about the data collected by these apps and their management. Contact tracing is of paramount importance when dealing with a pandemic, as it allows for rapid identification of cases based on the information collected from infected individuals about other individuals they may have had recent contact with. Advances in digital technology have enabled devices such as mobile phones to be used in the contract-tracing process. However, there is a potential risk of users' personal information and sensitive data being stolen should hackers be in the near vicinity of these devices. Thus, there is a need to develop privacy-preserving apps. Meanwhile, privacy policies that outline the risk associated with the use of contact-tracing apps are needed, in formats that are easily readable and comprehensible by the public. To our knowledge, no previous study has examined the readability of privacy policies of contact-tracings apps. Therefore, we performed a readability analysis to evaluate the comprehensibility of privacy policies of 7 contact-tracing apps currently in use. The contents of the privacy policies of these apps were assessed for readability using Readability Test Tool, a free web-based reliability calculator, which computes scores based on a number of statistics (ie, word count and the number of complex words) and indices (ie, Flesch Reading Ease, Flesch-Kincaid Reading Grade Level, Gunning Fog Index, and Simplified Measure of Gobbledygook index). Our analysis revealed that explanations used in the privacy policies of these apps require a reading grade between 7 and 14, which is considerably higher than the reading ability of the average individual. We believe that improving the readability of privacy policies of apps could be potentially reassuring for users and may help facilitate the increased use of such apps.

(J Med Internet Res 2020;22(12):e21572) doi:10.2196/21572

KEYWORDS

COVID-19; smartphone apps; contact tracing; privacy policy; readability; app; privacy; surveillance

Contact tracing is of paramount importance when dealing with a pandemic such as COVID-19. It allows for the rapid identification of cases based on the information collected from infected individuals about their recent contact with other individuals [1]. Additionally, contact-tracing systems allow for the collection of further information about these contacts, in order to help minimize the spread of the disease [2]. Different contact tracing apps rely on different technologies, including GPS, Bluetooth, and millimeter-wave communication technologies. Conventionally, public health care workers can assist an infected patient to map out individuals with whom they might have been in close contact recently, and these individuals are then informed of their susceptibility to the infection. Thus, contact tracing enables the identification of potential cases and allows for the follow-up and rapid quarantining of susceptible individuals [1]. With advances in digital technology, devices such as a mobile phone can now be employed in the contact-tracing process. For instance, a recent article in *Nature* discusses 3 apps developed to rapidly identify contacts of patients with COVID-19, including an Australia-based app called COVIDSafe and similar apps being used in Germany and Egypt [3].

Apps that enable contact tracing are instrumental in response to a public health emergency, but there have been concerns about the data they collect and how they are managed. Although there are potential benefits of using these apps, there are also ongoing concerns. For instance, in a recent commentary, Sharma et al [4] outlined the existing apps for COVID-19 contact tracing and concerns about data privacy. Another article in *Nature* cautioned against the accuracy of such contact-tracing apps and highlighted how these apps might render individuals susceptible to security breaches, given that most of these apps tap on Bluetooth functionality, potentially compromising the exchange of information [5]. This is an inherent risk that personal information and other sensitive personal data might be stolen if hackers happen to be in the vicinity of these devices [5]. A mixed-methods study in Norway analyzed the personal dataflows and the contents of privacy policies of 21 popular, free-to-use Android mobile apps [6]. Their results showed that 19 of the 21 apps studied transmitted personal data to about 600 different primary and third-party domains that were associated with tech companies in the United States. They also found that some apps tracked and shared data by default even when the app was not in use. The terms of use of some of these apps did not inform the users about the data sharing.

This study highlights critical ethical issues of data protection, security, and privacy of data collated by smartphone apps [1] and the consequent need to develop privacy-preserving smartphone apps [7]. A scoping review of the privacy assessments of mobile health apps reported that the evaluation criteria used in studies have been heterogeneous and lacked objectivity [8]. This further emphasizes the need for a common evaluation tool to ensure that personal health data privacy is safeguarded. It has also been suggested that a “healthcare fiduciary” be developed to enhance international regulatory frameworks to increase data protection security [9].

While we await the development of such privacy-preserving apps, privacy policies outlining the risks associated with the use of contact-tracing apps are needed, in a format that can be easily read and comprehended by the public. Readability of policy terms can be evaluated using validated tools that assess the complexity of the vocabulary and syntax, as well as the presentation of the content [10]. In other areas of health care, researchers have started to critique the readability of privacy policies. For instance, Robillard et al [10] focused on the availability and readability of privacy-related content of mental health apps and reported that most apps they studied did not include terms of agreement or a privacy policy. On the other hand, among the apps that had such policies in place, a reading ability more advanced than secondary education was required to comprehend the information. In relation to COVID-19, Basch et al [11] examined the information available on the internet and found that the readability levels required to comprehend the information exceeded that of the average American. The fact that higher-than-average readability levels are required to comprehend web-based information implies that the available information cannot be disambiguated, which might result in increased panic among the app users [11].

Given this situation, we performed a readability analysis of the privacy policies of 7 contact-tracing apps, namely COVIDSafe (used in Australia) [12], BeAware (used in Bahrain) [13], CoronApp (used in Colombia) [14], GH COVID-19 Tracker (used in Ghana) [15], Rakning C-19 (used in Iceland) [16], NZ COVID Tracer (used in New Zealand) [17], and TraceTogether (used in Singapore) [18]. As previously highlighted by Basch et al [11], the provision of timely information, in a format that could be comprehended easily, would help individuals understand important information relevant to the pandemic and, in turn, allay any anxieties. A readability analysis of privacy policies is timely and pertinent, given the considerable number of contact-tracing apps now available and government agencies’ enforcement that individuals download and use these apps. As a result, individuals are now more likely to examine the privacy policies of the apps they use, to understand what data is being shared and how their personal information is being protected. Any difficulty in comprehending the information contained within these privacy policies could result in a reluctance to download and use such apps.

Readability statistics of the privacy policies of the identified apps were computed using Readability Test Tool, a web-based reliability calculator [19]. This free resource computes the word count, Flesch Reading Ease, Flesch-Kincaid Reading Grade Level, Gunning Fog Index, Simplified Measure of Gobbledygook (SMOG) index, and the number of complex words [20]. For this evaluation, we used well-validated methods, based on previous studies that have examined readability [21,22]. The Flesch Reading Ease test evaluates the length of sentences and the number of polysyllabic words to determine the overall readability score; the score ranges from 0 to 100, with a higher score suggesting that the text is easy to read. The Flesh-Kincaid Reading Grade Level test evaluates the mean sentence and word length to compute reading complexity of the text; the score ranges from 1 to 12, corresponding to the US educational school grades, with scores higher than 12 indicative of college-level education and domain-specific experts. The Gunning Fog Index estimates the number of years of formal education required for an individual to understand the text on the first reading; the score ranges from 0 to 19+ and is representative of the readability level of the document. A Gunning Fog score of 0-6 is indicative of low literacy, a score of 7 or 8 is indicative of junior high school–level literacy, a score of 9-12 is indicative of high school–level literacy, a score of 13-16 is indicative of college-level literacy, a score of 17 or 18 is indicative of graduate-level literacy, and a score ≥19 suggests higher professional–level qualifications [23]. The SMOG index estimates the years of education needed to understand a piece of writing, by evaluating 10 sentences from the beginning, middle, and end of the document. The number of syllables in each section is then totaled and converted to a grade-level score [20]. Table 1 shows the readability scores for each of the 7 apps studied.

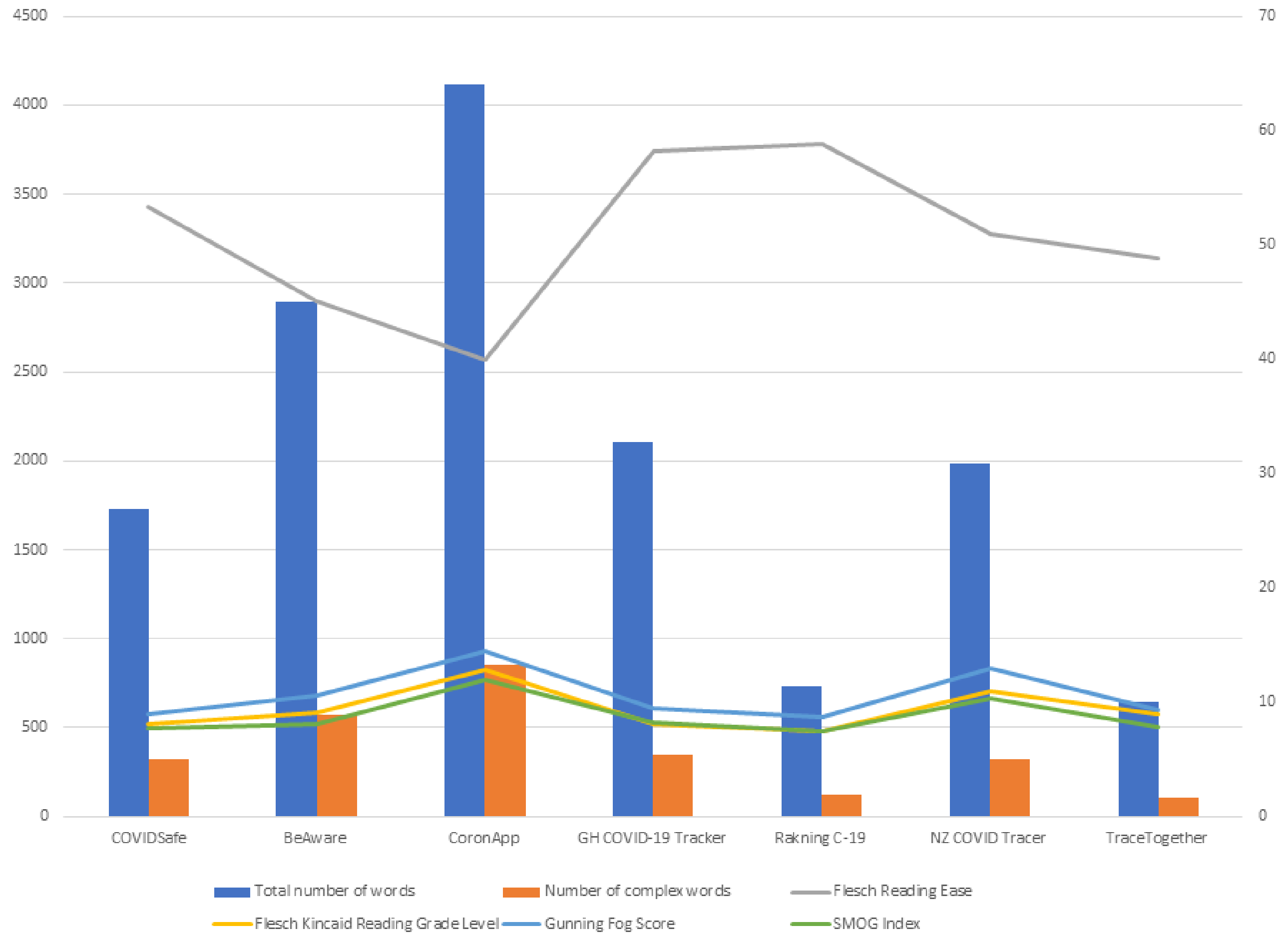
Table 1. Readability scores for the privacy policies of different COVID-19 contact-tracing apps analyzed in this study.

App name and description[a]	Readability scores					
	Total number of words, n	Complex-words, %	Flesch Kincaid Reading Ease	Flesch Kincaid Grade Level	Gunning Fog Score	SMOG index
COVIDSafe [12] • Developed by the Australian Government Department of health • Uses Bluetooth technology to record any contact one may have had with other users • Close contact information is securely stored on the phone, and it can be uploaded and used with the user's consent	1727	18.8	53.3	8.1	8.9	7.7
BeAware [13] • Helps contain COVID-19 spread by advancing contact-tracing efforts • Uses location data shared by users to alert individuals if they were in proximity with an active infected case • Tracks the movement of quarantine cases • Provides updates on COVID-19 developments and latest recommendations issued by health authorities	2893	19.9	45.1	9.1	10.6	8.1
CoronApp [14] • Official app by the Government of Chile to prevent COVID-19 spread • Allows self-assessment of symptoms • Allows user to receive notifications from the Ministry of Health and report high-risk behaviors	4119	20.7	39.9	12.8	14.5	12
GH COVID-19 Tracker [15] • Helps individuals assess & self-report symptoms and check risk of infection for COVID-19 • Includes the following features: check risk of infection, users near you, self-quarantine management, updates, and event management	2110	16.6	58.2	8.1	9.5	8.2
Rakning C-19 [16] • Official app by the Icelandic Government to help mitigate the COVID-19 pandemic in Iceland • Collects location information via GPS from the phone and stores data locally on the device • Assists in contact tracing	736	16.4	58.8	7.5	8.7	7.5
NZ COVID Tracer [17] • Official contact-tracing mobile app by the New Zealand Ministry of Health • Assists in contact tracing through the creation of a private digital diary of places the user has visited • Provides alerts if the user has checked into a place at the same time as someone with COVID-19	1990	16.3	50.9	11	13	10.3
TraceTogether [18] • Supports Singapore's efforts to fight the spread of COVID-19 through community-driven contact tracing • Notifies individuals if they have been exposed to COVID-19 through close contacts • Provides the latest guidance from the Singapore Ministry of Health • Uses Bluetooth, with the data being stored securely on the phone	645	16.7	48.8	9	9.3	7.8

[a] Each description of the apps have been summarized based on the original app descriptions listed on the app stores.

Users of contact-tracing apps must be aware that the apps gather a lot of their personal data, some from self-reporting and some via sensors in their smartphone devices. Moreover, our findings suggest that the existing explanations in the privacy policies of these apps require a reading level between 7 and 14, which far exceeds many people's reading ability. Apps like CoronApp [14] and NZ COVID Tracer [17] required the highest-grade level of comprehension (Figure 1), followed by BeAware [13], TraceTogether [18], GH COVID-19 Tracker [15], COVIDSafe [12], and Rakning C-19 [16], listed in order of decreasing readability ease. For example, in the United States, the average reading level is between grades 7 and 8 [24]. For the information to be accessible and achieve maximum impact among the general population, it should be written at a level no higher than grade 6 [22]. Hence, currently, the privacy policies of all the 7 apps analyzed in this study are considered "very difficult" to read and comprehend for the majority of individuals. In their analysis of the readability of online websites on COVID-19, Basch at al [11] highlighted how heightened levels of anxiety about the pandemic might further impair the understanding and interpretation of information, thus exacerbating fear.

Figure 1. Overview of readability scores for each COVID-19 contact-tracing app evaluated.



With more countries now exiting lockdowns, the use of contact-tracing apps will become more commonplace. While we await improvements to existing apps through the use of more secured technologies, the public must have access to readable terms of agreement or privacy policies to be aware of how their data are being collected, stored, and used. Improving the readability of privacy policies could be reassuring and could facilitate the adoption and eventual impact of these apps. Our study has highlighted COVID-19 contact-tracing apps with privacy policies that are readily understandable by the general public. Government agencies need to recognize this and to adapt their privacy policies accordingly, to ensure that every user can readily understand how their data are being stored and shared by the app. At a macro level, health care ministries and organizations could consider enhancing current regulatory frameworks to increase data protection security [9]. This may cause a trickle-down effect to app developers and companies and to the users, for safeguarding personal data.

Several research implications arise from our study findings. We concur with the suggestions by Bahadori et al [23] that researchers could undertake a number of measures to improve app readability. Users are also occasionally involved in the conceptualization of the app and in user testing. With the increase in participatory research, potential users could perhaps be involved in the cocreation and drafting of the privacy policies for such apps. Academics and developers are encouraged to consider the average reading level of the population when they are drafting these policies. As highlighted by Bahadori et al [23], an effective way to do so is to reduce the length of the sentence and target towards a reading level of grade 6. For continued monitoring of user experiences, they also recommend determining whether readability needs to be improved on an ongoing basis. As these areas develop, an objective evaluation tool should also be developed to assess whether sufficient measures have been taken to safeguard the data of mobile app users. By increasing the level of trust that users have in how an app uses their data, more users will be confident of using these apps. This will bode well as health care research drives into the age of big data to improve health care services for everyone.

Acknowledgments

MWZ is supported by a grant under the Singapore Ministry of Health’s National Medical Research Council (grant number NMRC/Fellowship/0048/2017) for PhD training. The funding source was not involved in any part of this project.

Authors' Contributions

MZ, AC, and HS jointly conceptualized the study. MZ and AC were involved in data extraction and verification of the extracted data. MZ worked on the first draft of the manuscript, and HS provided guidance and suggested revisions. MZ and AC amended the second draft of the manuscript. HS provided critical updates to the final manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
2. LI J, Guo X. Global deployment mappings and challenges of contact-tracing apps for COVID-19. SSRN Preprint posted online May 24, 2020. [doi: 10.2139/ssrn.3609516]
3. Show evidence that apps for COVID-19 contact-tracing are secure and effective. Nature 2020 Apr;580(7805):563. [doi: 10.1038/d41586-020-01264-1] [Medline: 32350479]
4. Sharma T, Bashir M. Use of apps in the COVID-19 response and the loss of privacy protection. Nat Med 2020 Aug;26(8):1165-1167. [doi: 10.1038/s41591-020-0928-y] [Medline: 32457443]
5. Zastrow M. Coronavirus contact-tracing apps: can they slow the spread of COVID-19? Nature Epub ahead of print posted online on May 19, 2020. [doi: 10.1038/d41586-020-01514-2] [Medline: 32433633]
6. Brandtzaeg PB, Pultier A, Moen GM. Losing control to data-hungry apps: a mixed-methods approach to mobile app privacy. Social Science Computer Review 2018 May 31;37(4):466-488. [doi: 10.1177/0894439318777706]
7. Yasaka T, Lehrich B, Sahyouni R. Peer-to-peer contact tracing: development of a privacy-preserving smartphone app. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
8. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy assessment in mobile health apps: scoping review. JMIR Mhealth Uhealth 2020 Jul 02;8(7):e18868 [FREE Full text] [doi: 10.2196/18868] [Medline: 32459640]
9. Galvin HK, DeMuro PR. Developments in privacy and data ownership in mobile health technologies, 2016-2019. Yearb Med Inform 2020 Aug;29(1):32-43 [FREE Full text] [doi: 10.1055/s-0040-1701987] [Medline: 32823298]
10. Robillard JM, Feng TL, Sporn AB, Lai J, Lo C, Ta M, et al. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. Internet Interv 2019 Sep;17:100243 [FREE Full text] [doi: 10.1016/j.invent.2019.100243] [Medline: 30949436]
11. Basch CH, Mohlman J, Hillyer GC, Garcia P. Public health communication in time of crisis: readability of on-line cOVID-19 information. Disaster Med Public Health Prep 2020 May 11:1-3 [FREE Full text] [doi: 10.1017/dmp.2020.151] [Medline: 32389144]
12. COVIDSafe App. Australian Government. URL: https://www.covidsafe.gov.au/ [accessed 2020-06-04]
13. BeAware Bahrain. Kingdom of Bahrain - eGovernment Apps Store. URL: https://apps.bahrain.bh/CMSWebApplication/action/ShowAppDetailsAction?selectedAppID=321&appLanguage=en [accessed 2020-06-04]
14. CoronApp. Government of Colombia. URL: https://coronaviruscolombia.gov.co/Covid19/aislamiento-saludable/coronapp.html [accessed 2020-06-04]
15. GH COVID-19 Tracker App. Ministry of Communications, Ghana. URL: https://ghcovid19.com/ [accessed 2020-06-04]
16. Rakning C-19 App. Google Play Store. URL: https://play.google.com/store/apps/details?id=is.landlaeknir.rakning&hl=en_SG [accessed 2020-06-04]
17. NZ COVID Tracer. Ministry of Health – Manatū Hauora. URL: https://tracing.covid19.govt.nz/ [accessed 2020-06-04]
18. TraceTogether, safer together. A Singapore Government Agency Website. URL: https://www.tracetogether.gov.sg/ [accessed 2020-06-04]

19. Readability test tool - Quick and easy way to test the readability of your work. WebFX. URL: https://www.webfx.com/tools/read-able/check.php?tab=Test+By+Url&uri=https%3A%2F%2Ffacetagr.com%2Fprivacy-policy%2F [accessed 2020-05-22]
20. Grabeel K, Russomanno J, Oelschlegel S, Tester E, Heidel R. Computerized versus hand-scored health literacy tools: a comparison of Simple Measure of Gobbledygook (SMOG) and Flesch-Kincaid in printed patient education materials. J Med Libr Assoc 2018 Jan;106(1):38-45 [FREE Full text] [doi: 10.5195/jmla.2018.262] [Medline: 29339932]
21. Fowler L, Gillard C, Morain S. Readability and accessibility of terms of service and privacy policies for menstruation-tracking smartphone applications. Health Promot Pract 2020 Sep;21(5):679-683. [doi: 10.1177/1524839919899924] [Medline: 32037887]
22. Ayyaswami V, Padmanabhan D, Crihalmeanu T, Thelmo F, Prabhu A, Magnani J. Mobile health applications for atrial fibrillation: a readability and quality assessment. Int J Cardiol 2019 Oct 15;293:288-293. [doi: 10.1016/j.ijcard.2019.07.026] [Medline: 31327518]
23. Bahadori S, Wainwright T, Ahmed O. Readability of information on smartphone apps for total hip replacement and total knee replacement surgery patients. J Patient Exp 2020 Jun;7(3):395-398 [FREE Full text] [doi: 10.1177/2374373519844266] [Medline: 32821800]
24. McKenzie J, Neiger B, Thackeray R. Planning, Implementing, and Evaluating Health Promotion Programs: A Primer. 7th Edition. New York, NY: Pearson; 2017.

Edited by G Eysenbach, R Kukafka; submitted 18.06.20; peer-reviewed by C Shen, J Li, M Adly, A Dawood, A McLean, A Azzam; comments to author 09.09.20; revised version received 25.09.20; accepted 08.10.20; published 03.12.20.

Please cite as:
Zhang M, Chow A, Smith H
COVID-19 Contact-Tracing Apps: Analysis of the Readability of Privacy Policies
J Med Internet Res 2020;22(12):e21572
URL: https://www.jmir.org/2020/12/e21572
doi:10.2196/21572
PMID:33170798

Viewpoint

The 21st Century Cures Act: A Competitive Apps Market and the Risk of Innovation Blocking

William J Gordon[1,2,3], MD, MBI; Kenneth D Mandl[3,4], MD, MPH

[1]Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States
[2]Mass General Brigham, Boston, MA, United States
[3]Harvard Medical School, Boston, MA, United States
[4]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States

Corresponding Author:
Kenneth D Mandl, MD, MPH
Computational Health Informatics Program
Boston Children's Hospital
300 Longwood Ave
Landmark 5506, Mail Stop BCH3187
Boston, MA,
United States
Phone: 1 6173554145
Email: Kenneth_Mandl@harvard.edu

Abstract

The 21st Century Cures Act and the recently published "final rule" define standardized methods for obtaining electronic copies of electronic health record (EHR) data through application programming interfaces. The rule is meant to create an ecosystem of reusable, substitutable apps that can be built once but run at any hospital system "without special effort." Yet, despite numerous provisions around information blocking in the final rule, there is concern that the business practices that govern EHR vendors and health care organizations in the United States could still stifle innovation. We describe potential app ecosystems that may form. We caution that misaligned incentives may result in anticompetitive behavior and purposefully limited functionality. Closed proprietary ecosystems may result, limiting the value derived from interoperability. The 21st Century Cures Act and final rule are an exciting step in the direction of improved interoperability. However, realizing the vision of a truly interoperable app ecosystem is not predetermined.

(J Med Internet Res 2020;22(12):e24824) doi:10.2196/24824

KEYWORDS

interoperability; application programming interfaces; SMART on FHIR; health information; patient records; digital infrastructure; digital; app ecosystem

Introduction

In May 2020, the Department of Health and Human Services and the Office of the National Coordinator for Health Information Technology (ONC) published a final rule implementing health information technology provisions of the 21st Century Cures Act [1]. The rule defines standardized methods for obtaining computable, electronic copies of electronic health record (EHR) data through an application programming interface (API). APIs enable granular, computable, immediate access to data and allow patients or provider organizations to connect an app directly to the EHR that exchanges data without an intermediary. The rule standardizes the Fast Healthcare Interoperability Resources (FHIR) data model and the Substitutable Medical Applications, Reusable Technologies (SMART) on FHIR standard [2,3], which specifies how users are authorized and how apps launch. The rule restricts providers and EHR vendors from "information blocking" — preventing the exchange of electronic health information.

Intense lobbying against the rule, preceding its release, by EHR vendors and hospital systems signals that control over data and revenue are both at stake. The rule not only enables patient and provider access to medical record data in computable formats but also shifts how they interact with the health care system, which entities control those interactions, and the underlying business models that could either threaten or accelerate interoperability's potential to improve care delivery.

Supporting an Interoperable App Economy

The rule is designed to foster a competitive ecosystem of substitutable (often third-party) apps that can be written once and connect to standardized health system data anywhere. The Cures Act requires an API that makes "all data elements" of a patient's record available "without special effort" [4]. Examples of apps that leverage such an API include Apple's Health Records on iPhone product, which enables patients to download computable copies of their medical records from hundreds of health systems [5]; a neonatal bilirubin management app that improves management [6]; and a commercially successful app, sold through multiple app stores, that dynamically generates patient-specific medication instructions [7]. Because apps access data using the SMART on FHIR API, they can be installed or deleted at any institution, independent of the underlying EHR, akin to smartphone apps. The "without special effort" clause should necessitate that the public API enable this capability without purchasing paywalled features from an EHR vendor. An interoperable apps–based economy could drive down costs, support improved public health surveillance and response [8], bring machine learning to the point of care [9], and accelerate genome-informed medicine [10,11].

Smartphone apps are distributed through app stores. The Apple App store and Google Play store, for example, are large "one-stop shopping" markets, enable turnkey installation, and offer customer review aggregation, technical and security reviews, and payment processing. As such, they are a unique and invaluable channel to deliver technology to an end user. However, innovators are charged as much as 30% of revenue, and Apple has found itself under intense scrutiny for controlling both the platform (the iPhone) and the marketplace (the App Store) [12]. Since EHR vendors have taken a similar tack [13], the final ONC rule wisely prohibits anticompetitive behaviors, including offering different service terms to similar apps, noncompete and exclusivity clauses, and intellectual property–transfer requirements. The major EHR vendors now have app store equivalents, like the Epic App Orchard, Cerner App Gallery, or AllScripts Application Store, as do new entrant companies constructing proprietary health care app ecosystems separate from the EHR vendors.

Risks of Misaligned Incentives and Anticompetitive Behavior

Despite information blocking provisions, business practices could stifle innovation and reduce choice. First of all, there is an issue of pricing and fees. Under the HITECH Act, the United States has already invested US $48 billion toward the promotion of EHR adoption. Because most EHR products are based on pre-internet software, the opportunity to layer on a modern infrastructure is essential for progress. The final ONC rule does not quantify permitted API fees, but allows for recouping fees "reasonably incurred" by the EHR vendor. How costs will be passed down is unclear. Health system leaders would be right to ask how much they should be expected to pay to get their own data out of EHR products they have already purchased. Patients too have already effectively paid for their data, through insurance premiums, taxes, or directly out of pocket. The Apple and Google app stores are proven to inspire developers to produce millions of apps. It is far from clear to us that the EHR marketplaces would lead to similar much-needed innovation in health care.

Secondly, we are concerned that EHR vendors may limit functionality and data availability across the public APIs and instead shift app connections to higher functioning proprietary APIs. The rule would permit an EHR vendor to profit from value-added services, as long as those services are not necessary for developing and deploying software that uses the API. It is extremely important that the public, regulated APIs give patients, providers, and innovators robust functionality. One issue is that the rule only requires a subset of data elements to be exposed through public APIs — the US Core Data for Interoperability (USCDI). Another is that the rule focuses on reading data from the EHR, rather than writing data back to the EHR. If the government does not expand the USCDI rapidly enough, or stalls on advancing a write capability, the proprietary APIs may outpace open, public, standardized APIs. An app written once would require a different version and set of agreements for each marketplace, analogous to a need to create a different version of a web page for every different web browser. Furthermore, EHR vendors may circumvent the spirit of the Cures Act and ONC rule by levying prohibitive revenue-sharing schemes for apps that access a modified version of the public API, charge for favorable placement in EHR-associated app stores, or limit important usability features, for example, requiring nonpreferred apps to frequently "log back in" to receive updated data. Health care organizations may be unaware of these innovation-blocking behaviors and powerless to stop them.

Thirdly, the ONC final rule applies to certified API developers — for now, these are predominantly EHR vendors. However, the rule may not apply to emerging secondary platforms that use the public API to move data into a proprietary system with a proprietary API, for example solutions built on iOS or Android. We could end up exchanging one closed ecosystem (EHRs) for another (secondary platforms), which will further segment the market, block innovation, and limit physician and patient choice.

Finally, assessing real-world implementation of the Cures Act will be challenging. Infractions of the "without special effort" provision could emerge through interpretations of the final rule, be hidden behind business contracts and nondisclosure agreements, or be promulgated through hesitancy to address EHR vendor business practices. Initial API usage has not been high [14,15], and much of it is attributable to Apple Health Records on iPhone alone. The good news is that the slow pace allows time to shape the unfolding ecosystem as the rule's provisions take effect over the next 2 years. Both manual and automated processes are needed to measure key provisions of the final rule [16] and assess whether the Cures Act has produced a robust apps economy where an app written once can run widely throughout health care.

Conclusion

While the Cures Act and ONC final rule place guardrails around information blocking, a truly interoperable plug-and-play app ecosystem is far from predetermined. There is ample room for "innovation blocking" even by vendors who are regulatorily compliant. Measurement of progress toward an open app ecosystem and additional regulation are needed to ensure return on the massive investment in national digital infrastructure.

Conflicts of Interest

WG reports receiving research funding from IBM and consulting fees from the Office of the National Coordinator, Department of Health and Human Services, United States. KM chairs the scientific advisory Board for Medal, Inc. Boston Children's Hospital receives corporate philanthropic support for KM's laboratory from SMART Advisory Committee members, which include the American Medical Association, the BMJ Group, Eli Lilly and Company, First Databank, Google Cloud, Hospital Corporation of America, Microsoft, Optum, Premier Inc, and Quest Diagnostics.

References

1. Department of Health and Human Services. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. Federal Register. 2020 May 01. URL: https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification [accessed 2020-11-28]
2. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep 17;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
3. Mandl KD, Mandel JC, Kohane IS. Driving Innovation in Health Systems through an Apps-Based Information Economy. Cell Syst 2015 Jul;1(1):8-13 [FREE Full text] [doi: 10.1016/j.cels.2015.05.001] [Medline: 26339683]
4. Public Law 114 - 255 - 21st Century Cures Act. U.S. Government Information. 2016 Dec 13. URL: https://www.govinfo.gov/app/details/PLAW-114publ255 [accessed 2020-11-30] [WebCite Cache ID 77b7kKmJx]
5. Gordon WJ, Bates DW, Fuchs D, Pappas J, Silacci S, Landman A. Comparing Characteristics of Patients Who Connect Their iPhones to an Electronic Health Records System Versus Patients Who Connect Without Personal Devices: Cohort Study. J Med Internet Res 2019 Aug 22;21(8):e14871 [FREE Full text] [doi: 10.2196/14871] [Medline: 31441430]
6. Kawamoto K, Kukhareva P, Shakib JH, Kramer H, Rodriguez S, Warner PB, et al. Association of an Electronic Health Record Add-on App for Neonatal Bilirubin Management With Physician Efficiency and Care Quality. JAMA Netw Open 2019 Nov 01;2(11):e1915343 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.15343] [Medline: 31730181]
7. Mandl KD, Gottlieb D, Ellis A. Beyond One-Off Integrations: A Commercial, Substitutable, Reusable, Standards-Based, Electronic Health Record-Connected App. J Med Internet Res 2019 Feb 01;21(2):e12902 [FREE Full text] [doi: 10.2196/12902] [Medline: 30707097]
8. Mishra NK, Duke J, Lenert L, Karki S. Public health reporting and outbreak response: synergies with evolving clinical standards for interoperability. J Am Med Inform Assoc 2020 Jul 01;27(7):1136-1138. [doi: 10.1093/jamia/ocaa059] [Medline: 32692844]
9. Gruendner J, Schwachhofer T, Sippl P, Wolf N, Erpenbeck M, Gulden C, et al. KETOS: Clinical decision support and machine learning as a service – A training and deployment platform based on Docker, OMOP-CDM, and FHIR Web Services. PLoS ONE 2019 Oct 3;14(10):e0223010 [FREE Full text] [doi: 10.1371/journal.pone.0223010] [Medline: 31581246]
10. Alterovitz G, Warner J, Zhang P, Chen Y, Ullman-Cullere M, Kreda D, et al. SMART on FHIR Genomics: facilitating standardized clinico-genomic apps. J Am Med Inform Assoc 2015 Nov;22(6):1173-1178. [doi: 10.1093/jamia/ocv045] [Medline: 26198304]
11. Ryu B, Shin S, Baek R, Kim J, Heo E, Kang I, et al. Clinical Genomic Sequencing Reports in Electronic Health Record Systems Based on International Standards: Implementation Study. J Med Internet Res 2020 Aug 10;22(8):e15040 [FREE Full text] [doi: 10.2196/15040] [Medline: 32773368]
12. Antitrust: Commission opens investigations into Apple's App Store rules. European Commission. 2020 Jun 16. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1073 [accessed 2020-11-28]
13. Allen A. Developers complain of high EHR fees for SMART apps. Politico. 2018 Aug 06. URL: https://www.politico.com/story/2019/09/23/developers-ehr-fees-smart-apps-721486 [accessed 2020-11-28]
14. Adler-Milstein J, Longhurst C. Assessment of Patient Use of a New Approach to Access Health Record Data Among 12 US Health Systems. JAMA Netw Open 2019 Aug 02;2(8):e199544 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.9544] [Medline: 31411708]
15. Gordon WJ, Patel V, Thornhill W, Bates DW, Landman A. Characteristics of Patients Using Patient-Facing Application Programming Interface Technology at a US Health Care System. JAMA Netw Open 2020 Oct 01;3(10):e2022408 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22408] [Medline: 33107918]

16. Savage M, Neinstein A, Adler-Milstein J. Measure The Impact Of The ONC's New Interoperability Rules Now. Health Affairs Blog. 2020 Jul 20. URL: https://www.healthaffairs.org/do/10.1377/hblog20200701.58142/full/ [accessed 2020-11-28]

Abbreviations

API: application programming interface
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resource
ONC: Office of the National Coordinator for Health Information Technology
SMART: Substitutable Medical Applications, Reusable Technologies
USCDI: US Core Data for Interoperability

Edited by G Eysenbach; submitted 07.10.20; peer-reviewed by N Sharma, J Lee; comments to author 28.10.20; revised version received 01.11.20; accepted 09.11.20; published 11.12.20.

Please cite as:
Gordon WJ, Mandl KD
The 21st Century Cures Act: A Competitive Apps Market and the Risk of Innovation Blocking
J Med Internet Res 2020;22(12):e24824
URL: http://www.jmir.org/2020/12/e24824/
doi:10.2196/24824
PMID:33306034

Viewpoint

Paradigm Shift Toward Digital Neuropsychology and High-Dimensional Neuropsychological Assessments: Review

Thomas Parsons[1], PhD; Tyler Duffield[2], PhD

[1]Computational Neuropsychology & Simulation, University of North Texas, Denton, TX, United States
[2]Oregon Health & Science University, Portland, OR, United States

Corresponding Author:
Thomas Parsons, PhD
Computational Neuropsychology & Simulation
University of North Texas
UNT Discovery Park, G Wing, 150 Learning Technologies 3940 N. Elm, G150
Denton, TX, 76207
United States
Phone: 1 940 565 2057
Fax: 1 940 565 2057
Email: thomas.parsons@unt.edu

Abstract

Neuropsychologists in the digital age have increasing access to emerging technologies. The National Institutes of Health (NIH) initiatives for behavioral and social sciences have emphasized these developing scientific and technological potentials (eg, novel sensors) for augmented characterization of neurocognitive, behavioral, affective, and social processes. Perhaps these innovative technologies will lead to a paradigm shift from disintegrated and data-poor behavioral science to cohesive and data-rich science that permits improved translation from bench to bedside. The 4 main advances influencing the scientific priorities of a recent NIH Office of Behavioral and Social Sciences Research strategic plan include the following: integration of neuroscience into behavioral and social sciences, transformational advances in measurement science, digital intervention platforms, and large-scale population cohorts and data integration. This paper reviews these opportunities for novel brain-behavior characterizations. Emphasis is placed on the increasing concern of neuropsychology with these topics and the need for development in these areas to maintain relevance as a scientific discipline and advance scientific developments. Furthermore, the effects of such advancements necessitate discussion and modification of training as well as ethical and legal mandates for neuropsychological research and praxes.

(J Med Internet Res 2020;22(12):e23777) doi:10.2196/23777

KEYWORDS

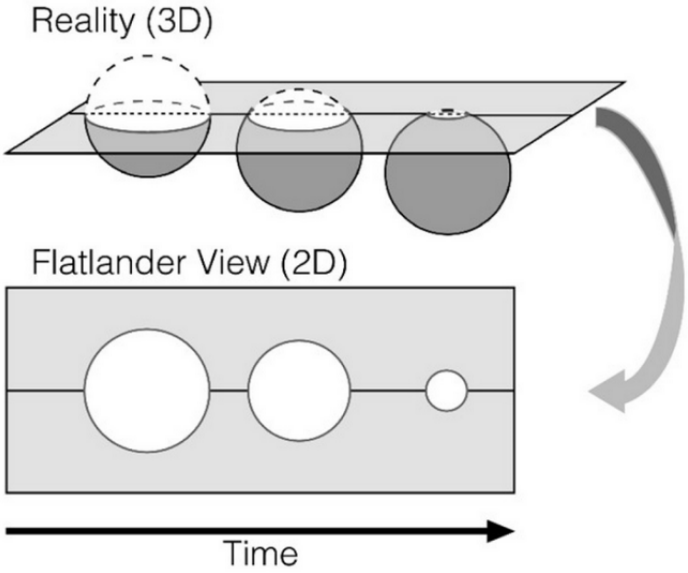
neuropsychology; technology; informatics; machine learning; big data; virtual reality; smartphone; mobile phone

Introduction

Clinical neuropsychologists have traditionally developed and validated parsimonious assessment tools using basic technologies (ie, pencil and paper protocols, general linear model). Advances have predominantly occurred in expanded normative standards throughout the history of this profession [1]. Although these low-dimensional tools are well-validated assessments of basic cognitive constructs, they have limited presentation (static 2D stimuli) and logging capabilities (which require manual logging of responses). Moreover, low-dimensional approaches limit their statistical modeling (typically linear) to combinations of features relative to a set of weights for predicting the value of criterion variables. Some neuropsychologists may argue that the parsimony offered by low-dimensional tools reflects the reality of a much higher-dimensional deficit. However, low-dimensional tools may offer diminished interpretations of complex phenomena.

The preference for low-dimensional tools is apparent in surveys of assessments used by neuropsychologists [2,3]. This conservativism has resulted in neuropsychological assessments that have hardly changed since the original scales were established in the early 1900s [4,5]. Low-dimensional neuropsychological assessment tools place the neuropsychologist on par with the 19th century literary work on the nature of perception and dimensionality. Specifically, the narrator, A Square, resides in Flatland with residents (Flatlanders) whose perception is limited to 2 dimensions. After a discussion with a *Stranger* (a sphere), A Square comes to understand how complex and high dimensional the world is. Unfortunately, A Square is jailed for holding and communicating heretical beliefs [6]. For neuropsychologists, low-dimensional technologies have led us to search for simplified explanations of complex phenomena, which limits our ability to develop, validate, interpret, and communicate useful models of human neuropsychology. Recently, cognitive psychologists have called this the *Flatland fallacy*. They contend that the Flatland fallacy can be surmounted by formalizing psychological theories as computational models that have the capacity to make precise predictions about cognition and/or behavior (Figure 1) [7].

Figure 1. A Square cannot perceive his world as anything other than 2-dimensional. Reprinted with permission.



These authors explain that in the limited perspective of the Flatlanders' view (bottom of the figure), a 3D object (sphere) seems to be fluctuating magnitudes (an expanding and reducing circle). However, the reality is that (top of the figure) this object is merely progressing through a lower - dimensional plane. The low-dimensional perspective leads to a false understanding of reality. Similarly, neuropsychologists may incorrectly determine that the low level of dimensionality correctly describes neuropsychological or psychological phenomena. In fact, they may be missing the complexity and high dimensionality of neuropsychological phenomena.

Cognitive psychologists also contend that unitary cognitive constructs such as *attention* are limited and prevent psychologists from deepening the understanding of complex, or high-dimensional, phenomena. First, theories of unitary cognitive constructs are based on circular reasoning. Complex phenomena such as the conception of *attention* are explained by presumptive attentional systems. Instead, psychologists should model the parallel, reciprocal, and iterative interactions between context and neural or functional processes. This would enhance the characterization of physically executed actions [8]. Similarly, the analytical approach to psychology is problematic because (1) an exhaustive definition is proposed (eg, attention), (2) assumed subfunctions are identified (eg, selective, sustained, or divided attention) with separable functional and neuronal processes (or dedicated systems), and (3) research concentrates on specific tasks that purport to measure the theoretical subfunctions rather than underlying processes required to execute an efficient behavior in a particular situation [9]. Commonalities between subfunctions and other constructs (eg, working memory) are often not empirically distinguishable and by no means imply that the underlying functional and neural processes are different or separable. These authors propose that rather than being rigidly adherent to prior cognitive conceptual frameworks, psychologists should model mechanisms and processes (sensory, motor, and cognitive) that are found in several complex behaviors. These behaviors may run in parallel or interact across stimulus properties, time, and goals to achieve an outcome.

How do neuropsychologists move from low-dimensional neuropsychology to high-dimensional neuropsychology? The National Institutes of Health (NIH) offers initiatives for neuropsychologists interested in higher-dimensional tools, including (1) integrating neuroscience into behavioral and social sciences, (2) transformative advances in measurement science, (3) digital intervention platforms, and (4) large-scale population cohorts and data integration [10]. Similarly, the NIH Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative seeks high-dimensional approaches to understand brain disorders (eg, Alzheimer disease, Parkinson disease, depression, and traumatic brain injury) and accelerating the technologies for high-dimensional modeling of how the brain records, processes, uses, stores, and retrieves vast quantities of information [11]. Neuropsychologists can enhance work conducted in NIH initiatives by offering interpretations of neuroscience findings based on clinical expertise.

After a brief consideration of the historic progression of neuropsychological assessment technologies, there is an elucidation of current NIH initiatives for the behavioral and social sciences as well as evaluations of current neuropsychological assessment technologies.

A Very Brief History of Neuropsychological Assessment Technologies

Neuropsychology has experienced a number of advances as it developed from a primarily qualitative practice to a more objective and evidence-based approach [12], with expanded normative standards [1], performance validity testing [13], and cross-cultural considerations [14]. Although these improvements have aided the investigation of neurocognitive functions, there are increasing discussions on the need to enhance the dimensionality of neuropsychological assessments and computational modeling [1,4,5,15-22].

The technological and theoretical development of neuropsychological assessment can be understood in terms of dimensional waves of technological adoption [5]. In Neuropsychology 1.0, neuropsychological assessments accentuate the development of low-dimensional and construct-driven (ie, simple stimulus presentations of stimuli to test abstract concepts like working memory) paper-and-pencil measures. In Neuropsychology 2.0, there is a technological move to automated administration, scoring, and in some instances the interpretation of low-dimensional stimulus presentations using computerized approaches (eg, NIH Toolbox and video teleconferencing) [23-26]. Concurrently, technological developments in neuroimaging have changed the role of neuropsychological assessments, from lesion localization to predictions about a patient's ability to perform activities of daily living. Finally, Neuropsychology 3.0 reflects contemporary advances in high-dimensional (dynamic and ecologically valid simulation, logging, and modeling of everyday activities) tools.

Some neuropsychologists are hesitant to move from low-dimensional to high-dimensional tools because computerized assessments may introduce errors and/or decrease the reliability of the assessment process by means of automation [27]. Although there have been improvements in computational power and security, developers of high-dimensional technologies need to take appropriate actions to ensure proper implementation [28]. Furthermore, normative efforts are ongoing for high-dimensional assessments, and continued validation of advanced platforms and novel data analytic approaches is needed.

Three decades ago, clinical psychologists were urged to adopt progressively available advanced technologies [29]. Concurrently, in the 1980s, neuropsychologists started discussing the potential of computerized neuropsychological assessments [30]. It was subsequently pointed out that when compared with progress in our everyday technologies, psychological assessments had barely progressed throughout the 20th century (eg, Wechsler scales) [31]. Technologies found in neuropsychological testing can be compared with now obsolete black-and-white televisions, vinyl records, rotary-dial telephones, and the first commercial computer made in the United States (ie, Universal Automatic Computer I). Assessment technologies need to progress in ideas, not just new measures [31].

In the late 1990s, it was discussed how neuropsychology lagged behind (in absolute terms and in comparison with) other clinical neurosciences. Clinical neuropsychologists continued to use many of the same tools that have been developed decades earlier. Moreover, new tests that were coming out were not conceptually or substantively better than the ones from decades earlier (eg, Wechsler scales). Dodrill [32] pointed to the fact that in the 1970s, there was little difference in the technological progress of neurology and neuropsychology. This changed with the advent of computerized tomographic scanning, and neuropsychologists were no longer consulted for lesion localization. Neuroimaging advances allowed neurologists to better understand and treat neurologic pathophysiology [33]. Dodrill [32] suggests that if technological developments in neurology had been as slow as that found in neuropsychology, then neurologists would be limited to pneumoencephalograms and radioisotope scans. These procedures are deemed primeval by current neuroradiological standards.

To get an idea of where neuropsychology is today, basic searches were performed on July 29, 2020, to tally the number of technology publications per discipline. The first search included a PubMed search with the search terms "computer" AND ("neuropsychology" OR "neurology" OR "neuroscience") from 1990 to 2019 (Figure 2). A second and third search using the terms "technology" and "neuroimaging" instead of "computer" revealed similar findings (Figures 3 and 4, respectively). Figures 2-4 show the number of publications by year that resulted from each of these 3 broad literature searches.

Figure 2. Proliferation of publications identified in the PubMed database over time. Search terms: "computer" by discipline (eg, "neuropsychology," "neurology," "neuroscience").

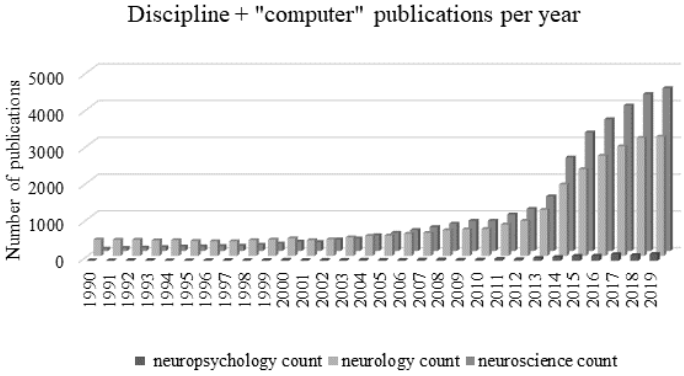


Figure 3. Proliferation of publications identified in the PubMed database over time. Search terms: “technology” by discipline (eg, “neuropsychology,” “neurology,” “neuroscience”).

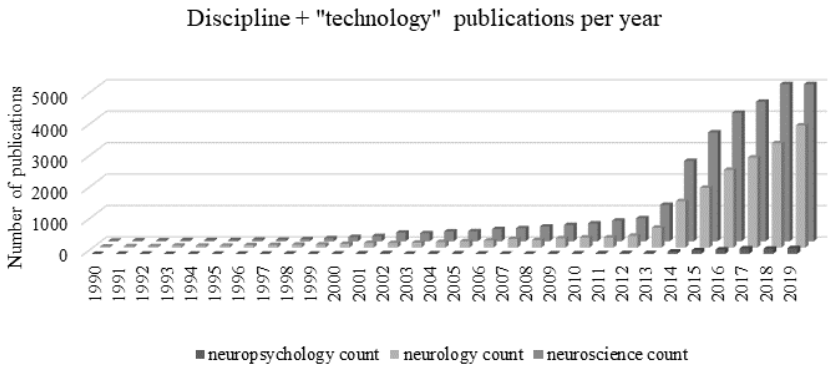
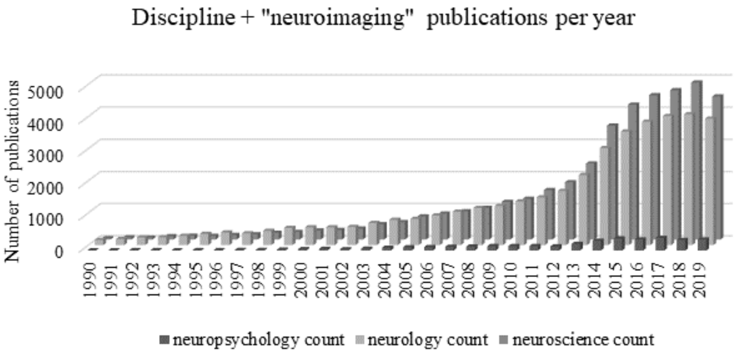


Figure 4. Proliferation of publications identified in the PubMed database over time. Search terms: “neuroimaging” by discipline (eg, “neuropsychology,” “neurology,” “neuroscience”).



Findings from these basic searches suggest that high-dimensional technologies have vastly greater representations in neurology and the neurosciences. The inclusion of technologies is very recently increasing in neuropsychology but is explicitly not keeping pace with other neurosciences. Similarly, a survey of rates of neuropsychologists using computerized instruments revealed that only 6% of the 693 neuropsychology assessments were computerized [3]. The average respondent reported that they rarely used computerized tests. An increased likelihood of computerized assessment use was apparent for early career neuropsychologists.

NIH’s Transformative Opportunities for the Behavioral and Social Sciences

Integrating Neuroscience Advances Into Clinical Neuropsychology

High-dimensional technologies such as functional neuroimaging provide real-time observations of brain function that challenge the validity of some low-dimensional paper-and-pencil technologies. Impairments following brain injury are rarely a single type of processing, and there is no one-to-one relationship between neuropsychological functions and brain structures and systems. Similar symptoms can arise from various injury types, and the same underlying injury can result in a variety of different symptoms. Although the integration of neuroimaging and neuropsychological methods has improved our understanding of brain functions, specific neuropsychological functions are typically associated with activation in multiple brain regions (distributed processing). Advances in methods and high-dimensional technologies offer promise for redefining previous understandings of cognitive functions (eg, elucidation of multiple types of processing speed) and introduction of novel (and complex and dynamic) cognitive functions [34].

Neuropsychologists are increasingly arguing for neuropsychological models established in terms of patients’ reciprocal relations with the environments in which they carry out activities of daily living [35-37]. Understanding the complex and dynamic interactions of persons involves the study of the brain’s processes in environmental and social systems. The increasing emphasis of clinical neuropsychology on ecological validity [38,39] and integration with social neuroscience [40,41] is limited by current low-dimensional paper-and-pencil assessments [42]. There is growing attention to the development of high-dimensional tools for assessing and modeling brain functions that include dynamic presentations of environmentally relevant stimuli [36,43,44]. Moving beyond the static or low-dimensional stimuli found in most traditional neuropsychological tests require neuropsychologists to find ways to update their technologies to reflect high-dimensional assessment approaches (eg, deep learning, mobile platforms, wearables, extended reality [XR], and the *Internet of Things* [IoT]).

Neuropsychologists have looked to factor analytic studies of neuropsychological test results to enhance understanding of the functional capacities of patients [45]. However, looking at the relations among responses to low-dimensional tasks that use static or 2D stimuli can constrain task performance and neural activity to abstract constructs (eg, working memory). Low-dimensional assessments bind mean neural population dynamics to a low-dimensional subspace and may limit the neuropsychologist’s assessment of the patient’s ability to perform everyday activities [46]. Furthermore, observation of low-dimensional neural signals may be an artifact of simple cognitive tasks. Standard paper-and-pencil (low-dimensional) tasks often involve basic responses to static or low-dimensional stimuli [47].

Computational neuroscience offers high-dimensional models of cognition via neural network–motivated connectionist models. This approach integrates neuroscience findings into high-dimensional models of the ways in which our brains execute cognitive functions. Leabra is a programing framework that has been used to integrate connectionist models of cognitive function. The result is a holistic architecture adept at producing precise predictions of a broad array of cognitive processes [48,49]. Computational models based on neuroscience findings allow for assessing a model’s sensitivity for capturing a neuropsychological construct and specificity of a given construct to other neuropsychological states and processes. Finally, computational models are shareable and extensible by other neuroscientists who want to extend previous work via iterative construct validation.

Adoption of Advances in Measurement Science to Neuropsychological Assessment

The NIH Office of Behavioral and Social Sciences Research (OBSSR) also emphasizes advances in measurement science and the move from low-dimensional data analytical approaches (typically linear combinations of features relative to a set of weights for criterion value prediction) to higher-dimensional data analytic approaches for evaluating change over time (eg, deep learning neural networks, machine learning). Clinical scientists are starting to adopt developments in deep learning and other computational modeling approaches [50]. Machine learning and deep learning have been applied successfully in various areas of artificial intelligence research: natural language processing, speech recognition, self-driving cars, and computer vision. For example, natural language processing–oriented computerized neuropsychological assessments have been developed to extract key features of linguistic complexity changes associated with progression in the Alzheimer disease spectrum [51]. High-dimensional data analytics can be applied to computerized adaptive testing (CAT) and computational models derived from large collaborative databases.

High-dimensional measurement protocols offer a clinical scientist with increased precision and granularity of data [10]. Technologically enhanced neuropsychological assessments (including high-dimensional virtual environments [VEs] with graphical models) surpass simple automations (computerized neuropsychological assessments) of low-dimensional paper-and-pencil tasks. Moreover, they allow neuropsychologists to present scenarios that require patients to actively choose among multiple subtasks. From higher-dimensional tasks, context-dependent computational models can be established that include latent context variables that can be extricated using nonlinear modeling.

A framework has been proposed that aims to elucidate probabilistic computations using graphical and statistical models of naturalistic behaviors. The probability distribution for high-dimensional (ecologically valid simulations of everyday activities) tasks is complex. As a result, the brain likely simplifies the high-dimensional stimuli by centering on significant interactions [47]. Neuropsychologists can develop probabilistic graphical models for proficient descriptions of complex statistical distributions that relate several interactions and/or conditional dependencies among neuropsychological variables.

Deep Learning for Higher-Dimensional Algorithms

Neuropsychologists can use deep learning algorithms that simulate the hierarchical structure of a person’s brain (healthy and damaged). Deep learning is a form of machine learning (ie, algorithms that learn from data to automatically perform tasks such as classification and prediction that can be nonlinear in nature) that processes data from lower dimensionality to increasingly higher dimensions. It is increasingly used to develop novel technologies, big data, and artificial intelligence [52]. Neuropsychologists can use deep learning to analyze studies with both traditional (low-dimensional paper and pencil) and high-dimensional simulation technologies (eg, virtual reality–based neuropsychological assessments, mixed reality, augmented reality). With deep learning, neuropsychologists could process the lower-dimensional data (paper-and-pencil tests). Next, they could move to increasingly higher-dimensional data (eg, from simulation technologies) and develop increasingly complex data-driven semantic concepts that are likely more representative of brain functioning than historical, theoretically based cognitive constructs (eg, working memory).

Probabilistic models and generative neural networks can be used to develop a unified framework for modeling neuropsychological functioning (nonclinical and clinical). Connectionist models such as these are understood to be a portion of the more general framework of probabilistic graphical models. Neuropsychological performances have been modeled as Bayesian computations (brain function expresses perceptions and actions as inferential processes). In this approach, neuropsychological deficits are false inferences arising from aberrant previous beliefs. Bayesian approaches can be used for computational phenotyping that uses graphical models implemented as stochastic processes that involve a randomly determined sequence of observations (each of which is considered as a sample of one element from a probability distribution) via generative neural networks [53]. Visual object recognition (eg, facial processing) can be used as an example. Selective lesions can be applied to computational models of visual object recognition to assess the impact of damage to various cortical regions (eg, early visual processing, extrastriate areas, anterior associative areas). New high-dimensional measures could be developed to assess visual agnosia and examine the appearance of category-specific deficits.

Deep learning architectures can also be used for modeling specific connection pathways in selective impairment. Stochastic decay (stochastic reduction of weight values that decreases responsivity to afferent signals) can be applied to synaptic strengths for examination of cognitive decline. Both global degradation of all network synapses and local degradation of inhibitory synapses from a given processing layer have been investigated. The findings revealed that although older participants accurately performed arithmetical tasks, they had impaired numerosity discrimination on trials requiring the inhibition of incongruent information. They also found that these results were related to poor inhibitory processes measured by standard neuropsychological assessments. The specific degradation of inhibitory processes resulted in a pattern closely resembling older participants' performance [54]. The addition of computational modeling for the development, validation, and application of neuropsychological assessments represents a high-dimensional approach for neuropsychologists.

CAT and Item Response Theory

The NIH Toolbox is a battery of computerized neuropsychological assessments that uses item response theory (IRT) and CAT. With IRT, the NIH Toolbox has an alternative to classical test theory as it moves beyond group-specific norms [55]. In IRT, the probability of an item response is modeled according to the respondent's position on the underlying construct of interest. This approach can be useful for providing item-level properties of each NIH Toolbox measure across the full range of each construct. Although neuropsychological measures tend to meet the reliability and validity requirements of classical test theory, the equivalence of item properties (eg, item difficulty and item discriminatory power) is often assumed across items. Consideration of item difficulty tends to be subsumed under independent variable manipulation (eg, cognitive load) to modify the marginal likelihood of correct responses in item subgroups. A limitation of this approach is that it does not match well with current item-level analyses found in neuroimaging assessments of brain activations following stimulus probes. For neuropsychological assessments to comport well with brain activation probes, item difficulty needs to be considered to avoid ceiling and floor effects in patient performances across clinical cohorts. IRT models offer the neuropsychologist both individual patient parameters and individual item characteristics that can be scaled along a shared latent dimension. Neuropsychological assessments would benefit from greater adoption of developments in IRT that emphasize the accuracy of individual items. Various IRT approaches have been applied as signal detection theory models that connect corresponding but discrete methods [56]. Combining IRT and signal detection delivers the measurement accuracy needed for robust modeling of item difficulty and examinee ability.

The NIH Toolbox CAT approach shortens testing time (by about half as long as low-dimensional paper-and-pencil measures). Through avoidance of floor or ceiling effects and concise item pools, CAT delivers equal (or greater) ability–level assessments [57,58]. Moreover, CAT offers enhanced efficiency, flexibility, and precision assessment of multiple domains of interest without adversely affecting participant burden. The application of IRT to CAT provides neuropsychologists with real-time assessment of item-level performance.

Function-Led Assessments Using High-Dimensional Simulations

Neuropsychologists are increasingly interested in developing assessments that assess the patients' real-world *functions* in a manner that generalizes to functional performance in everyday activities [38]. A *function-led approach* to neuropsychological assessments involves starting with directly observable everyday behaviors and proceeding backward to observe how a sequence of actions leads to a given behavior. Furthermore, a function-led approach examines how that behavior is disrupted. For example, a patient may have difficulty multitasking while using a global positioning system to navigate a simulated neighborhood in a driving simulator. High-dimensional technologies can be used to present dynamic and interactive stimuli in a 3D environment that includes automatic logging and computational modeling (eg, head movements, eye tracking, response latencies, patterns, etc) of a patient's performance in everyday activities. High-dimensional neuropsychology tools are being developed and validated to simulate everyday functions (rather than abstract cognitive constructs) [5,41].

Given the drawbacks to experiments conducted in real-life settings (time consuming, require transportation, involve consent from local businesses, costly to use or build physical mock-ups, and difficult to replicate or standardize across settings) and difficulty in maintaining systematic control of real-world stimulus challenges, high-dimensional and function-led XR environments are being used by neuropsychologists.

Low-dimensional (paper-and-pencil and computer automated) neuropsychological tools only indirectly assess the patient's ability to perform everyday activities [39,59]. VEs offer potential aids in enhancing the dimensionality and ecological validity of neuropsychological assessments through enhanced computational capacities for administration efficiency, stimulus presentation, automated logging of responses, and data analytic processing. Given the precise stimulus presentation and control of dynamic or high-dimensional perceptual stimuli, VEs offer neuropsychological assessments with enhanced ecological validity [5,60-62]. High-dimensional immersive VEs move beyond low-dimensional paper-and-pencil tests with static stimulus presentations in sterile testing rooms to simulated environments that replicate the distractions, stressors, and/or demands found in the real world.

Data Monitoring With High-Dimensional Technologies

Using passive data monitoring from everyday technologies (eg, smartphones, IoT), clinical scientists can collect real-time cognitive performance throughout the course of a day [10]. Each patient has a digital footprint that transpires from consistent use of everyday technologies. Coupling technologies with developments in measurement science allows for novel approaches to capture cognitions, affects, and behaviors [63]. Rapid progress in sensor technologies has led to objective and effective measures of behavioral performance, psychophysiology, and environmental contexts [64]. For example, machine learning has been employed to extract features from passive monitoring of mobile phone use. When comparing these features with performance on the psychomotor vigilance task, it was found that alertness deviations as small as 11% could be detected [65].

Another example of enhanced data monitoring can be found in the increased granularity in performance assessments and digital logging tools used in the Framingham Heart Study [66]. New developments in digital logging of verbal responses to cognitive stimuli allow for automated algorithms that can extract new language features (eg, speaker turn taking, speaking rate, hesitations, pitch, number of words, vocabulary). These features offer promise for predicting incident cognitive impairment [67]. Furthermore, low-dimensional pencils and pens can be upgraded with high-dimensional digital pens with associated software designed to measure pen positioning 75 times per second. Digital pens have a spatial resolution of ±0.002 inches. For example, digital pens are being used by neuropsychologists for assessing clock drawing performance [68,69]. Minute drawing elements such as pen strokes (eg, clock face, hand, digit) can be logged with greater than 84% accuracy [70]. The sensitivity of these high-dimensional technologies to minute drawing elements, decision-making latencies, and graphomotor characteristics may offer promise to greatly enhance lower-dimensional hand scoring of the Boston Process Approach. A review of the Boston Process Approach and neuropsychology-based technologies has been available recently [71].

Digital Intervention Platforms

Another transformative opportunity from the NIH OBSSR is the application of high-dimensional technologies to interventions [10]. Progress in neurocognitive rehabilitation has been enhanced by neuroimaging of plasticity of the brain. Similarly, a notable increase can be found in the use of noninvasive brain stimulation approaches that leverage neural plasticity for rehabilitation [72]. Neuropsychologists interested in rehabilitation emphasize the promotion of brain plasticity by increasing a patient's capacity for performing everyday activities. The resource and labor intensiveness of interventions and the resulting limitations (reach, scalability, and duration) found in real-world assessment environments require interventions to be personalized at the start, adapted throughout treatment, and operationalized into coded databases for fidelity [10].

Smart Environment Technologies

Smart environments integrate and incorporate several high-dimensional capabilities (eg, function-led evaluation, passive data monitoring, deep learning, etc) to provide both assessment and intervention. Using smart environments, neuropsychologists can discreetly monitor a patient's everyday activities for changes in clinical status (eg, mobility patterns can predict neurocognitive status). Moreover, automatic interventions can be provided in real-world settings [73-77]. Smart environments use machine learning algorithms (eg, naïve Bayes, Markov, conditional random fields, and dynamic Bayes networks) to model, recognize, and monitor large amounts of labeled training data [78,79]. Activity aware prompting is used to assist in the elevation of independent living. Results from studies using prompting technologies reveal growth in independent activity engagement by patients with neurocognitive impairment [80,81].

VE Technologies

Smart virtual reality environments simulate real-world scenarios and offer refined stimulus delivery for interventions [60,82,83]. Using VEs, neuropsychologists can present and control stimuli across various sensory modalities (eg, visual, auditory, olfactory, haptic, and kinesthetic). There is an increasing number of validated VEs that can be used for assessment and intervention: virtual apartments [84], grocery stores [85], libraries [86], classrooms [87-90], driving [91], cities [92,93], and military environments [94,95]. In addition to the use of novel measurement science for more efficient assessments using behavioral performances, real-time psychophysiological data (eg, eye gaze) can also be used to adapt assessment and intervention environments for a more individualized approach using factors such as emotional reactivity and ongoing skill development [5,64].

Smartphones and Other Digital Technologies

Current NIH initiatives for the behavioral and social sciences contend that intervention technologies need to move from short-term assessment and rehabilitation interventions (low-dimensional assessments and treatments that may limit *maintenance* of behavioral response and change) to high-dimensional approaches that use novel technologies (eg, smartphones) to extend treatment duration to improve behavioral maintenance [10]. Mobile technologies offer neuropsychologists higher-dimensional interventions that extend into patients' everyday activities by logging, monitoring, prompting, and skill building between treatment sessions. One version of this involves ecological momentary assessments and interventions, as patients perform activities of daily living [96-98]. Ecological momentary assessments and interventions using digital devices offer large streams of continuous data [99,100]. Advances in computational modeling offer distinctive prospects for real-time behavioral interventions in ecological contexts [101-104]. As with any new tool, neuropsychologists need to develop and validate measures and interventions.

Large-Scale Population Cohorts, Data Integration, and Cognitive Ontologies

The NIH OBSSR strategic plan is also interested in big data, data analytics, and data integration techniques for developing collaborative knowledge bases [10]. Integrating neuropsychological data into large collaborative knowledge bases will allow neuropsychologists to either formalize cognitive ontologies or abandon cognitive ontologies for phylogenetically refined functional and neuronal processes that underlie all complex behaviors or more simplistically traditional neuropsychological tasks [8,32,38,105]. Formal designations of distinct sensory, motor, and cognitive entities can be established in terms of parallel, reciprocal, hierarchical, and/or spatiotemporal relations [7,18].

Consistent with critiques from cognitive psychology [8], a limitation of neuropsychological data integration is that low-dimensional neuropsychological assessments are made up of hypothetical interdimensional constructs inferred from research findings [19,38]. Evidence for poor test specificity is apparent in median correlations for common neuropsychological tests. It has been found that although the median correlation within domain groupings on a neuropsychological test was 0.52, the median correlation between groupings was 0.44 [32]. Therefore, the tests are not unambiguously domain specific. The median correlations should be notably higher within groupings and lower between groupings. A recent meta-analysis of relationships between the Wisconsin Card Sorting Test (WCST) and the Weschler Adult Intelligence Scale (WAIS) found a robust relationship between WCST performance and WAIS indices [106]. This is interesting because the WAIS was recently found to be the test most often administered by neuropsychologists and the WCST was the fifth most often administered [2]. Interestingly, the meta-analysis found that WCST scores were associated in comparable strength with both verbal and nonverbal domains from Wechsler Adult Intelligence Scale tests. Another issue is that there is considerable variation in some neuropsychological tests of the same domain (eg, various measures of go or no-go performance) [107]. The shared variance of tests of supposedly differing domains and the lack of consistency in tests of the same domain may decrease the capacity for accurate data integration.

Compounding this issue is the fact that current diagnostic frameworks found in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM) and the World Health Organization's International Classification of Diseases (ICD) are dependent on presenting signs and symptoms. Moreover, they do not align with findings from genetics and clinical neuroscience [108,109].

Ontologies are formal specifications of entities found in a domain and their relations. An ontology contains designations of separate entities along with a specification of ontological relations among entities with representations via spatiotemporal (eg, *preceded-by* or *contained-within*) or hierarchical relations (eg, *is-a* or *part-of*). This provision of an objective, concise, common, and controlled vocabulary facilitates communication among domains. Neuropsychological assessment lags behind other clinical sciences in the development of formal ontologies [18,19].

As such, neuropsychologists have moved beyond the diagnostic taxonomies found in the DSM and ICD. These diagnostic taxonomies are not sufficient for biomarker research because they do not reflect relevant neurocognitive and behavioral systems. Instead, neuropsychologists interested in developing a common vocabulary for ontologies and collaborative knowledge bases should adopt the US National Institute of Mental Health's Research Domain Criteria (RDoC) project. The RDoC aims to establish a classification system for mental disorders based on neuroscience and behavioral research [108,110,111].

Conclusions

Neuropsychologists interested in high-dimensional technologies have embraced the following NIH initiatives to advance scientific developments: (1) integration of neuroscience into behavioral and social sciences, (2) transformative advances in measurement science, (3) digital intervention platforms, and (4) large-scale population cohorts and data integration. Evidence that progress is occurring in neuropsychology exists; however, more work needs to be done. Much of this work involves adoption, development, and validation of novel technologies. Similarly, there is a need for a classification system (based on neuroscience and psychology research) that moves beyond low-dimensional emphases on unitary cognitive constructs specific to a purported functional or neuronal system. A high-dimensional classification instead embraces testable hypotheses of how an observed phenomenon is produced from fundamental underlying mechanisms or processes, the dynamics of those processes (eg, reciprocal, hierarchal, iterative), and the multiple functional or neuronal systems involved in several complex behaviors [8,34]. In more basic terms, neuropsychologists should theorize with verbs instead of nouns to serve scientific progress. Only then can neuropsychologists integrate data to develop meaningful ontologies and collaborative knowledge bases of high-dimensional neuropsychological phenomena. Computational modeling has great promise for achieving this endeavor.

High-dimensional neuropsychology requires substantial reform in the way the profession conducts training. High-dimensional training should be added to current trainings that emphasize primarily (in some programs it may be solely) low-dimensional neuropsychological tests (eg, paper-and-pencil tests) and methods (limited introduction to general linear modeling). Increased emphasis should be placed on technical skill development with high-dimensional technologies and data-driven inferential reasoning. Curricula in neuropsychology programs should be expanded to adapt to the recent technological advances that have led to exponential growth in the other sciences. This would require reimagining training in clinical psychology programs. If neuropsychologists of the future are to work with large collaborative knowledge bases and perform complicated computational modeling of big data, then they need at least basic training in areas traditionally associated with computer science (eg, computer programing) and informatics (algorithms and databases). As such, their basic statistical training would need to be enhanced to include data manipulation, predictive model generation, machine learning, natural language processing, graph theory, and visualization. Increased emphasis on training basic technical and computational skills will improve the ability of future neuropsychologists to participate in science.

A final note is the need for training in neuroethics. Neuroethics has been distinguished into 2 branches: (1) ethics of neuroscience—neuroethics as applied ethical reflection on the practices and technologies found in the neurosciences—and (2) neuroscience of ethics—what neuroscience can reveal about the nature of morality and morally relevant topics [112]. Neuroethics are important for the NIH BRAIN initiative. The

NIH BRAIN project aims to examine the ways in which dynamic patterns of neural activity are transformed into cognition, emotion, perception, and action in health and disease [113]. The BRAIN initiative promotes the use of powerful new tools and technologies: (1) technologies for monitoring neural circuit activity and (2) technologies that enable the modulation of neural circuits [114]. As expected, the ethical concerns related to the medical and nonmedical use of neurotechnologies by neuropsychologists are profound. Neuroethics for neurotechnologies include a combination of principlist, deontological, and consequential ethical approaches to answer ethical quandaries [115,116]. Training in neuroethics and the ethical use of high-dimensional technologies will allow neuropsychologists to provide enhanced care for their patients.

Conflicts of Interest

None declared.

References

1. Casaletto KB, Heaton RK. Neuropsychological assessment: past and future. J Int Neuropsychol Soc 2017 Dec 4;23(9-10):778-790. [doi: 10.1017/s1355617717001060]
2. Rabin LA, Paolillo E, Barr WB. Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: a follow-up survey of INS and NAN members. Arch Clin Neuropsychol 2016 May;31(3):206-230. [doi: 10.1093/arclin/acw007] [Medline: 26984127]
3. Rabin LA, Spadaccini AT, Brodale DL, Grant KS, Elbulok-Charcape MM, Barr WB. Utilization rates of computerized tests and test batteries among clinical neuropsychologists in the United States and Canada. Profess Psychol: Research Pract 2014 Oct;45(5):368-377. [doi: 10.1037/a0037987]
4. Miller J, Barr W. The technology crisis in neuropsychology. Arch Clin Neuropsychol 2017 Aug 1;32(5):541-554. [doi: 10.1093/arclin/acx050] [Medline: 28541383]
5. Parsons T. Clinical Neuropsychology and Technology: What’s New and How We Can Use It. New York, USA: Springer; 2016.
6. Abbott E. Flatland: A Romance of Many Dimensions. New York, USA: Dover Publications; 1952.
7. Jolly E, Chang LJ. The flatland fallacy: moving beyond low-dimensional thinking. Top Cogn Sci 2019 Apr;11(2):433-454 [FREE Full text] [doi: 10.1111/tops.12404] [Medline: 30576066]
8. Hommel B, Chapman CS, Cisek P, Neyedli HF, Song J, Welsh TN. No one knows what attention is. Atten Percept Psychophys 2019 Oct;81(7):2288-2303 [FREE Full text] [doi: 10.3758/s13414-019-01846-w] [Medline: 31489566]
9. Krauzlis RJ, Bollimunta A, Arcizet F, Wang L. Attention as an effect not a cause. Trends Cogn Sci 2014 Sep;18(9):457-464 [FREE Full text] [doi: 10.1016/j.tics.2014.05.008] [Medline: 24953964]
10. Collins F, Riley W. NIH's transformative opportunities for the behavioral and social sciences. Sci Transl Med 2016 Nov 23;8(366):366ed14 [FREE Full text] [doi: 10.1126/scitranslmed.aai9374] [Medline: 27881821]
11. Insel TR, Landis SC, Collins FS. Research priorities. The NIH BRAIN initiative. Science 2013 May 10;340(6133):687-688 [FREE Full text] [doi: 10.1126/science.1239276] [Medline: 23661744]
12. Duff K. Evidence-based indicators of neuropsychological change in the individual patient: relevant concepts and methods. Arch Clin Neuropsychol 2012 May;27(3):248-261 [FREE Full text] [doi: 10.1093/arclin/acr120] [Medline: 22382384]
13. Greher MR, Wodushek TR. Performance validity testing in neuropsychology. J Psychiat Pract 2017;23(2):134-140. [doi: 10.1097/pra.0000000000000218]
14. Olson K, Jacobson K. Cross-cultural considerations in pediatric neuropsychology: a review and call to attention. Appl Neuropsychol Child 2015;4(3):166-177. [doi: 10.1080/21622965.2013.830258] [Medline: 25117200]
15. Beaumont J. Introduction to Neuropsychology. New York, USA: Guilford Press; 2008.
16. Benton A, Sivan A. Clinical neuropsychology: a brief history. Dis Mon 2007 Mar;53(3):142-147. [doi: 10.1016/j.disamonth.2007.04.003] [Medline: 17544643]
17. Bilder RM, Reise SP. Neuropsychological tests of the future: how do we get there from here? Clin Neuropsychol 2019 Feb;33(2):220-245 [FREE Full text] [doi: 10.1080/13854046.2018.1521993] [Medline: 30422045]
18. Jagaroo V. Neuroinformatics for Neuropsychology. New York, USA: Springer; 2009.
19. Jagaroo V, Santangelo S. Advancing Psychiatry and Neuropsychology. New York, USA: Springer; 2017.
20. Kane R, Parsons T. The Role of Technology in Clinical Neuropsychology. Oxford, UK: Oxford University Press; 2017.
21. Parsons TD, Duffield T. National Institutes of Health initiatives for advancing scientific developments in clinical neuropsychology. Clin Neuropsychol 2019 Feb;33(2):246-270. [doi: 10.1080/13854046.2018.1523465] [Medline: 30760117]
22. Ponsford J. International growth of neuropsychology. Neuropsychology 2017 Nov;31(8):921-933. [doi: 10.1037/neu0000415] [Medline: 29376670]
23. Munro Cullum C, Hynan L, Grosch M, Parikh M, Weiner M. Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. J Int Neuropsychol Soc 2014 Oct 24;20(10):1028-1033. [doi: 10.1017/s1355617714000873]
24. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH toolbox for assessment of neurological and behavioral function. Neurology 2013 Mar 11;80(Issue 11, Supplement 3):S2-S6. [doi: 10.1212/wnl.0b013e3182872e5f]

25. Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioural function: the NIH toolbox. Lancet Neurol 2010 Feb;9(2):138-139. [doi: 10.1016/s1474-4422(09)70335-7]
26. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Slotkin J, et al. The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: validation in an adult sample. J Int Neuropsychol Soc 2014 Jun 24;20(6):567-578. [doi: 10.1017/s1355617714000320]
27. Cernich AN, Brennana DM, Barker LM, Bleiberg J. Sources of error in computerized neuropsychological assessment. Arch Clin Neuropsychol 2007 Feb;22(Suppl 1):S39-S48. [doi: 10.1016/j.acn.2006.10.004] [Medline: 17097851]
28. Parsons TD, McMahan T, Kane R. Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. Clin Neuropsychol 2018 Jan;32(1):16-41. [doi: 10.1080/13854046.2017.1337932] [Medline: 28590154]
29. Meehl P, Butcher F. Foreward. In: Psychological Assessment. New York, USA: Basic Books; 1987:xv-xvi.
30. Adams KM, Kvale VI, Keegan JF. Relative accuracy of three automated systems for neuropsychological interpretation. J Clin Neuropsychol 1984 Nov;6(4):413-431. [doi: 10.1080/01688638408401232] [Medline: 6548760]
31. Sternberg RJ. Intelligence and lifelong learning: what's new and how can we use it? Am Psychol 1997 Oct;52(10):1134-1139. [doi: 10.1037/0003-066x.52.10.1134]
32. Dodrill CB. Myths of neuropsychology. Clin Neuropsychol 1997 Feb;11(1):1-17. [doi: 10.1080/13854049708407025]
33. Cendes F, Theodore W, Brinkmann B, Sulc V, Cascino G. Neuroimaging of epilepsy. In: Handbook of Clinical Neurology. New York, USA: Elsevier; 2016:985-1014.
34. Price CJ. The evolution of cognitive models: from neuropsychology to neuroimaging and back. Cortex 2018 Oct;107:37-49 [FREE Full text] [doi: 10.1016/j.cortex.2017.12.020] [Medline: 29373117]
35. Bigler ED. Systems biology, neuroimaging, neuropsychology, neuroconnectivity and traumatic brain injury. Front Syst Neurosci 2016;10:55 [FREE Full text] [doi: 10.3389/fnsys.2016.00055] [Medline: 27555810]
36. Genon S, Reid A, Langner R, Amunts K, Eickhoff SB. How to characterize the function of a brain region. Trends Cogn Sci 2018 Apr;22(4):350-364 [FREE Full text] [doi: 10.1016/j.tics.2018.01.010] [Medline: 29501326]
37. Bigler E. Evidence-based integration of clinical neuroimaging findings in neuropsychology. Neuropsychol 2017:183.
38. Burgess PW, Alderman N, Forbes C, Costello A, Coates L, Dawsom DR, et al. The case for the development and use of 'ecologically valid' measures of executive function in experimental and clinical neuropsychology. J Int Neuropsychol Soc 2006 Mar 22;12(2):194-209. [doi: 10.1017/s1355617706060310]
39. Chaytor N, Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. Neuropsychol Rev 2003 Dec;13(4):181-197. [doi: 10.1023/b:nerv.0000009483.91468.fb]
40. Beauchamp MH. Neuropsychology's social landscape: common ground with social neuroscience. Neuropsychology 2017 Nov;31(8):981-1002. [doi: 10.1037/neu0000395] [Medline: 29376673]
41. Parsons TD, Carlew AR, Magtoto J, Stonecipher K. The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. Neuropsychol Rehabil 2017 Jul;27(5):777-807. [doi: 10.1080/09602011.2015.1109524] [Medline: 26558491]
42. Parsons T, Gaggioli A, Riva G. Virtual reality for research in social neuroscience. Brain Sci 2017 Apr 16;7(4):- [FREE Full text] [doi: 10.3390/brainsci7040042] [Medline: 28420150]
43. Barkley R. Executive functions. In: What They Are, How They Work, and Why They Evolved. New York, USA: Guilford Press; 2012.
44. Zaki J, Ochsner K. The need for a cognitive neuroscience of naturalistic social cognition. Ann N Y Acad Sci 2009 Jun;1167:16-30 [FREE Full text] [doi: 10.1111/j.1749-6632.2009.04601.x] [Medline: 19580548]
45. Larrabee GJ. Flexible vs fixed batteries in forensic neuropsychological assessment: reply to Bigler and Hom. Arch Clin Neuropsychol 2008 Jan 1;23(7-8):763-776. [doi: 10.1016/j.acn.2008.09.004]
46. Gao P, Ganguli S. On simplicity and complexity in the brave new world of large-scale neuroscience. Curr Opin Neurobiol 2015 Jun;32:148-155. [doi: 10.1016/j.conb.2015.04.003] [Medline: 25932978]
47. Pitkow X, Angelaki DE. Inference in the brain: statistics flowing in redundant population codes. Neuron 2017 Jun 7;94(5):943-953 [FREE Full text] [doi: 10.1016/j.neuron.2017.05.028] [Medline: 28595050]
48. O'Reilly R, Hazy T, Herd S. The leabra cognitive architecture. In: The Oxford Handbook of Cognitive Science. Oxford, UK: Oxford university Press; 2015:91-116.
49. O'Reilly R, Hazy T, Herd S. The leabra cognitive architecture: How to play 20 principles with nature. Oxford HandBook Congn Sci 2016;91:91-116. [doi: 10.1093/oxfordhb/9780199842193.013.8]
50. Menke A. PGPM 2018 Nov;Volume 11:211-212. [doi: 10.2147/pgpm.s146110]
51. Fang C, Janwattanapong P, Martin H. Computerized Neuropsychological Assessment in Mild Cognitive Impairment Based on Natural Language Processing-oriented Feature Extraction. In: International Conference on Bioinformatics and Biomedicine. 2017 Presented at: BIBM'17; June 1-5, 2017; New York, USA. [doi: 10.1109/bibm.2017.8217706]
52. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. J Big Data 2015 Feb 24;2(1):1-10. [doi: 10.1186/s40537-014-0007-7]
53. Testolin A, Zorzi M. Probabilistic models and generative neural networks: towards an unified framework for modeling normal and impaired neurocognitive functions. Front Comput Neurosci 2016 Jul 13;10:1-10. [doi: 10.3389/fncom.2016.00073]

54. Cappelletti M, Didino D, Stoianov I, Zorzi M. Number skills are maintained in healthy ageing. Cogn Psychol 2014 Mar;69:25-45. [doi: 10.1016/j.cogpsych.2013.11.004] [Medline: 24423632]
55. Thomas ML. The value of item response theory in clinical assessment: a review. Assessment 2011 Sep;18(3):291-307. [doi: 10.1177/1073191110374797] [Medline: 20644081]
56. Thomas ML, Brown GG, Gur RC, Moore TM, Patt VM, Risbrough VB, et al. A signal detection-item response theory model for evaluating neuropsychological measures. J Clin Exp Neuropsychol 2018 Oct;40(8):745-760 [FREE Full text] [doi: 10.1080/13803395.2018.1427699] [Medline: 29402152]
57. Gibbons RD, Weiss DJ, Kupfer DJ, Frank E, Fagiolini A, Grochocinski VJ, et al. Using computerized adaptive testing to reduce the burden of mental health assessment. Psychiatr Serv 2008 Apr;59(4):361-368 [FREE Full text] [doi: 10.1176/ps.2008.59.4.361] [Medline: 18378832]
58. Weiss DJ. Computerized adaptive testing for effective and efficient measurement in counseling and education. Measure Eval Counsel Dev 2017 Oct 19;37(2):70-84. [doi: 10.1080/07481756.2004.11909751]
59. Manchester D, Priestley N, Jackson H. The assessment of executive functions: coming out of the office. Brain Inj 2004 Nov;18(11):1067-1081. [doi: 10.1080/02699050410001672387] [Medline: 15545205]
60. Bohil CJ, Alicea B, Biocca FA. Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 2011 Nov 3;12(12):752-762. [doi: 10.1038/nrn3122] [Medline: 22048061]
61. Franzen M, Wilhelm K. Conceptual foundations of ecological validity in neuropsychological assessment. In: Ecological Validity of Neuropsychological Testing. New York, USA: Gr Press/St Lucie Press Inc; 1996:91-112.
62. Parsons TD. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front Hum Neurosci 2015;9:660 [FREE Full text] [doi: 10.3389/fnhum.2015.00660] [Medline: 26696869]
63. Reece AG, Danforth CM. Erratum to: Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017 Sep 5;6(1):1-12. [doi: 10.1140/epjds/s13688-017-0118-4]
64. Wade J, Zhang L, Bian D, Fan J, Swanson A, Weitlauf A, et al. A gaze-contingent adaptive virtual reality driving environment for intervention in individuals with autism spectrum disorders. ACM Trans Interact Intell Syst 2016 May 5;6(1):1-23. [doi: 10.1145/2892636]
65. Abdullah S, Murnane E, Matthews M. Cognitive Rhythms: Unobtrusive and Continuous Sensing of Alertness Using a Mobile Phone. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: ACM'16; April 13-15, 2016; New York, USA. [doi: 10.1145/2971648.2971712]
66. Au R, Piers RJ, Devine S. How technology is reshaping cognitive assessment: lessons from the Framingham heart study. Neuropsychology 2017 Nov;31(8):846-861. [doi: 10.1037/neu0000411]
67. Alhanai T, Au R, Glass J. Spoken Language Biomarkers for Detecting Cognitive Impairment. In: IEEE Automatic Speech Recognition and Understanding Workshop. 2017 Presented at: ASRU'17; May 3-6, 2017; New York, USA. [doi: 10.1109/asru.2017.8268965]
68. Piers RJ, Devlin KN, Ning B, Liu Y, Wasserman B, Massaro JM, et al. Age and graphomotor decision making assessed with the digital clock drawing test: the Framingham heart study. J Alzheimer's Dis 2017 Nov 7;60(4):1611-1620. [doi: 10.3233/jad-170444]
69. Dion C, Arias F, Amini S, Davis R, Penney D, Libon DJ, et al. Cognitive correlates of digital clock drawing metrics in older adults with and without mild cognitive impairment. J Alzheimer's Dis 2020 May 5;75(1):73-83. [doi: 10.3233/jad-191089]
70. Souillard-Mandar W, Davis R, Rudin C, Au R, Libon DJ, Swenson R, et al. Learning classification models of cognitive conditions from subtle behaviors in the digital Clock Drawing Test. Mach Learn 2015 Oct 20;102(3):393-441. [doi: 10.1007/s10994-015-5529-5]
71. Diaz-Orueta U, Blanco-Campal A, Lamar M, Libon DJ, Burke T. Marrying past and present neuropsychology: is the future of the process-based approach technology-based? Front Psychol 2020 Mar 6;11:1-8. [doi: 10.3389/fpsyg.2020.00361]
72. Crosson B, Hampstead B, Krishnamurthy L, Krishnamurthy V, McGregor KM, Nocera JR, et al. Advances in neurocognitive rehabilitation research from 1992 to 2017: the ascension of neural plasticity. Neuropsychology 2017 Nov;31(8):900-920. [doi: 10.1037/neu0000396]
73. Alberdi A, Aztiria A, Basarab A. On the early diagnosis of Alzheimer's disease from multimodal signals: a survey. Artif Intell Med 2016 Jul;71:1-29. [doi: 10.1016/j.artmed.2016.06.003]
74. Cook DJ, Schmitter-Edgecombe M. Assessing the quality of activities in a smart environment. Methods Inf Med 2018 Jan 20;48(5):480-485. [doi: 10.3414/me0592]
75. Dawadi PN, Cook DJ, Schmitter-Edgecombe M. Automated cognitive health assessment from smart home-based behavior data. IEEE J Biomed Health Inform 2016 Jul;20(4):1188-1194. [doi: 10.1109/jbhi.2015.2445754]
76. Galambos C, Skubic M, Wang S, Rantz M. Management of dementia and depression utilizing in-home passive sensor data. Gerontechnology 2013;11(3):457. [doi: 10.4017/gt.2013.11.3.004.00]
77. Schmitter-Edgecombe M, Cook D, Weakley A, Dawadi P. Using smart environment technologies to monitor and assess everyday functioning and deliver real-time intervention. In: The Role of Technology in Clinical Neuropsychology. Oxford, UK: Oxford University Press; 217:293-303.

78. Alberdi Aramendi A, Weakley A, Aztiria Goenaga A, Schmitter-Edgecombe M, Cook DJ. Automatic assessment of functional health decline in older adults based on smart home data. J Biomed Inform 2018 May;81:119-130. [doi: 10.1016/j.jbi.2018.03.009]
79. Singla G, Cook DJ, Schmitter-Edgecombe M. Recognizing independent and joint activities among multiple residents in smart environments. J Ambient Intell Human Comput 2009 Dec 1;1(1):57-63. [doi: 10.1007/s12652-009-0007-1]
80. Boger J, Mihailidis A. The future of intelligent assistive technologies for cognition: devices under development to support independent living and aging-with-choice. NeuroRehabilitation 2011 May 6;28(3):271-280. [doi: 10.3233/nre-2011-0655]
81. Boll S, Heuten W, Meyer EM, Meis M. Development of a multimodal reminder system for older persons in their residential home. Inform Health Soc Care 2010 Dec 6;35(3-4):104-124. [doi: 10.3109/17538157.2010.528651]
82. Wilson BA. ‘Cutting edge’ developments in neuropsychological rehabilitation and possible future directions. Brain Impairment 2012 Feb 21;12(1):33-42. [doi: 10.1375/brim.12.1.33]
83. Wilson B. Neuropsychological Rehabilitation: State of the Science. London, UK: Sage Publications; 2013.
84. Henry M, Joyal CC, Nolin P. Development and initial assessment of a new paradigm for assessing cognitive and motor inhibition: The bimodal virtual-reality Stroop. J Neurosci Method 2012 Sep;210(2):125-131. [doi: 10.1016/j.jneumeth.2012.07.025]
85. Parsons TD, Barnett M. Validity of a newly developed measure of memory: feasibility study of the virtual environment grocery store. J Alzheimer's Dis 2017 Aug 14;59(4):1227-1235. [doi: 10.3233/jad-170295]
86. Renison B, Ponsford J, Testa R, Richardson B, Brownfield K. The ecological and construct validity of a newly developed measure of executive function: the virtual library task. J Int Neuropsychol Soc 2012 Feb 20;18(03):440-450. [doi: 10.1017/s1355617711001883]
87. Iriarte Y, Diaz-Orueta U, Cueto E, Irazustabarrena P, Banterla F, Climent G. AULA—advanced virtual reality tool for the assessment of attention. J Atten Disord 2012 Dec 12;20(6):542-568. [doi: 10.1177/1087054712465335]
88. Parsons TD, Carlew AR. Bimodal virtual reality stroop for assessing distractor inhibition in autism spectrum disorders. J Autism Dev Disord 2015 Nov 27;46(4):1255-1267. [doi: 10.1007/s10803-015-2663-7]
89. Parsons TD, Bowerly T, Buckwalter JG, Rizzo AA. A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. Child Neuropsychology 2007 Jun 11;13(4):363-381. [doi: 10.1080/13825580600943473]
90. Parsons TD, Duffield T, Asbee J. A comparison of virtual reality classroom continuous performance tests to traditional continuous performance tests in delineating ADHD: a meta-analysis. Neuropsychol Rev 2019 Jun 3;29(3):338-356. [doi: 10.1007/s11065-019-09407-6]
91. Schultheis MT, Rebimbas J, Mourant R, Millis SR. Examining the usability of a virtual reality driving simulator. Assistive Technology 2007 Mar 31;19(1):1-10. [doi: 10.1080/10400435.2007.10131860]
92. Plancher G, Barra J, Orriols E, Piolino P. The influence of action on episodic memory: a virtual reality study. Q J Experi Psychol 2018 Jan;66(5):895-909. [doi: 10.1080/17470218.2012.722657]
93. Plancher G, Gyselinck V, Piolino P. The integration of realistic episodic memories relies on different working memory processes: evidence from virtual navigation. Front Psychol 2018 Jan 30;9:1-10. [doi: 10.3389/fpsyg.2018.00047]
94. Armstrong CM, Reger GM, Edwards J, Rizzo AA, Courtney CG, Parsons TD. Validity of the virtual reality stroop task (VRST) in active duty military. J Clin Experim Neuropsychol 2013 Feb;35(2):113-123. [doi: 10.1080/13803395.2012.740002]
95. Parsons TD, Courtney CG. An initial validation of the virtual reality paced auditory serial addition test in a college sample. J Neurosci Method 2014 Jan;222:15-23. [doi: 10.1016/j.jneumeth.2013.10.006]
96. Bylsma LM, Taylor-Clift A, Rottenberg J. Emotional reactivity to daily events in major and minor depression. J Abnormal Psychol 2011;120(1):155-167. [doi: 10.1037/a0021662]
97. Cook PF, McElwain CJ, Bradley-Springer LA. Feasibility of a daily electronic survey to study prevention behavior with HIV-infected individuals. Res. Nurs. Health 2010 Apr 27;33(3):221-234. [doi: 10.1002/nur.20381]
98. Runyan JD, Steenbergh TA, Bainbridge C, Daugherty DA, Oke L, Fry BN. A smartphone ecological momentary assessment/intervention 'app' for collecting real-time data and promoting self-awareness. PLoS ONE 2013 Aug 14;8(8):e71325. [doi: 10.1371/journal.pone.0071325]
99. Berke EM, Choudhury T, Ali S, Rabbi M. Objective measurement of sociability and activity: mobile sensing in the community. Ann Family Med 2011 Jul 11;9(4):344-350. [doi: 10.1370/afm.1266]
100. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Behav Med Pract Policy Res 2015 May 28;5(3):335-346. [doi: 10.1007/s13142-015-0324-1]
101. Nilsen WJ, Pavel M. Moving behavioral theories into the 21st century: technological advancements for improving quality of life. IEEE Pulse 2013 Sep;4(5):25-28. [doi: 10.1109/mpul.2013.2271682]
102. Rivera DE, Jimison HB. Systems modeling of behavior change: two illustrations from optimized interventions for improved health outcomes. IEEE Pulse 2013 Nov;4(6):41-47. [doi: 10.1109/mpul.2013.2279621]
103. Saranummi N, Spruijt-Metz D, Intille SS, Korhone I, Nilsen WJ, Pavel M. Moving the science of behavior change into the 21st century: novel solutions to prevent disease and promote health. IEEE Pulse 2013 Sep;4(5):22-24. [doi: 10.1109/mpul.2013.2271680]

104. Spring B, Gotsis M, Paiva A, Spruijt-Metz D. Healthy apps: mobile devices for continuous monitoring and intervention. IEEE Pulse 2013 Nov;4(6):34-40. [doi: 10.1109/mpul.2013.2279620]
105. Cisek P. Resynthesizing behavior through phylogenetic refinement. Atten Percept Psychophys 2019 Jun 3;81(7):2265-2287. [doi: 10.3758/s13414-019-01760-1]
106. Kopp B, Maldonado N, Scheffels JF, Hendel M, Lange F. A meta-analysis of relationships between measures of Wisconsin card sorting and intelligence. Brain Sciences 2019 Nov 29;9(12):349. [doi: 10.3390/brainsci9120349]
107. Sabb FW, Bearden CE, Glahn DC, Parker DS, Freimer N, Bilder RM. A collaborative knowledge base for cognitive phenomics. Mol Psychiatry 2008 Jan 8;13(4):350-360. [doi: 10.1038/sj.mp.4002124]
108. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med 2013 May 14;11(1):1-10. [doi: 10.1186/1741-7015-11-126]
109. Insel TR. Faulty circuits. Sci Am 2010 Apr;302(4):44-51. [doi: 10.1038/scientificamerican0410-44]
110. Insel TR, Scolnick EM. Cure therapeutics and strategic prevention: raising the bar for mental health research. Mol Psychiatry 2005 Dec 15;11(1):11-17. [doi: 10.1038/sj.mp.4001777]
111. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 2010 Jul;167(7):748-751. [doi: 10.1176/appi.ajp.2010.09091379] [Medline: 20595427]
112. Roskies A. Neuroethics for the new millenium. Neuron 2002 Jul;35(1):21-23. [doi: 10.1016/s0896-6273(02)00763-8]
113. Greely H, Ramos K, Grady C. Neuroethics in the age of brain projects. Neuron 2016 Nov;92(3):637-641. [doi: 10.1016/j.neuron.2016.10.048]
114. Ramos KM, Rommelfanger KS, Greely HT, Koroshetz WJ. Neuroethics and the NIH BRAIN initiative. J Responsible Innov 2017 May 5;5(1):122-130. [doi: 10.1080/23299460.2017.1319035]
115. Parsons T. Ethical Challenges in Digital Psychology and Cyberpsychology. Cambridge, USA: Cambridge University Press; 2019.
116. Farah M. An ethics toolbox for neurotechnology. Neuron 2015 Apr;86(1):34-37. [doi: 10.1016/j.neuron.2015.03.038]

Abbreviations

BRAIN: Brain Research through Advancing Innovative Neurotechnologies
CAT: computerized adaptive testing
DSM: Diagnostic and Statistical Manual of Mental Disorders
ICD: International Classification of Diseases
IoT: Internet of Things
IRT: item response theory
NIH: National Institutes of Health
OBSSR: Office of Behavioral and Social Sciences Research
RDoC: Research Domain Criteria
VE: virtual environment
WAIS: Weschler Adult Intelligence Scale
WCST: Wisconsin Card Sorting Test
XR: extended reality

Edited by G Eysenbach; submitted 22.08.20; peer-reviewed by U Diaz-Orueta, W Zhang; comments to author 13.10.20; revised version received 26.10.20; accepted 28.10.20; published 16.12.20.

Please cite as:
Parsons T, Duffield T
Paradigm Shift Toward Digital Neuropsychology and High-Dimensional Neuropsychological Assessments: Review
J Med Internet Res 2020;22(12):e23777
URL: http://www.jmir.org/2020/12/e23777/
doi:10.2196/23777
PMID:33325829

Viewpoint

Implications of the COVID-19 Pandemic for the Development and Update of Clinical Practice Guidelines: Viewpoint

Theresa Steeb[1], MPH; Markus Follmann[2], MPH, MD; Ralf Matthias Hagen[3], MD; Carola Berking[1], MD; Markus Vincent Heppt[1], MSc, MHBA, MD

[1]Department of Dermatology, Universitätsklinikum Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany

[2]German Cancer Society, Berlin, Germany

[3]Department of Microbiology and Hospital Hygiene, Bundeswehr Central Hospital, Koblenz, Germany

Corresponding Author:
Theresa Steeb, MPH
Department of Dermatology
Universitätsklinikum Erlangen
Friedrich-Alexander-University Erlangen-Nürnberg
Ulmenweg 18
Erlangen, 91054
Germany
Phone: 49 091318545852
Email: Theresa.Steeb@uk-erlangen.de

Abstract

Following the rapid spread of a new type of coronavirus (SARS-CoV-2), nearly all countries have introduced temporary restrictions affecting daily life, with “social distancing” as a key intervention for slowing the spread of the virus. Despite the pandemic, the development or actualization of medical guidelines, especially in the rapidly changing field of oncology, needs to be continued to provide up-to-date evidence- and consensus-based recommendations for shared decision making and maintaining the treatment quality for patients. In this viewpoint, we describe the potential strengths and limitations of online conferences for medical guideline development. This viewpoint will assist guideline developers in evaluating whether online conferences are an appropriate tool for their guideline conference and audience.

(J Med Internet Res 2020;22(12):e20064) doi:10.2196/20064

KEYWORDS

practice guideline; consensus development conference; guideline; videoconferencing; clinical practice; COVID-19; pandemic; public health; policy; decision making; online conference

The world currently faces an outbreak of a new type of coronavirus (SARS-CoV-2), which began spreading in China at the end of 2019 and has now led to a worldwide pandemic [1]. The syndrome caused by SARS-CoV-2 (COVID-19) manifests clinically diversely, with some patients having an asymptomatic disease course while others suffer from pneumonia or life-threatening acute respiratory distress syndrome, which may lead to multisystem organ failure or death [2,3]. Since health care systems worldwide struggle with an unexpected large number of cases that need to be hospitalized, this pandemic represents a global public health threat of international concern [4].

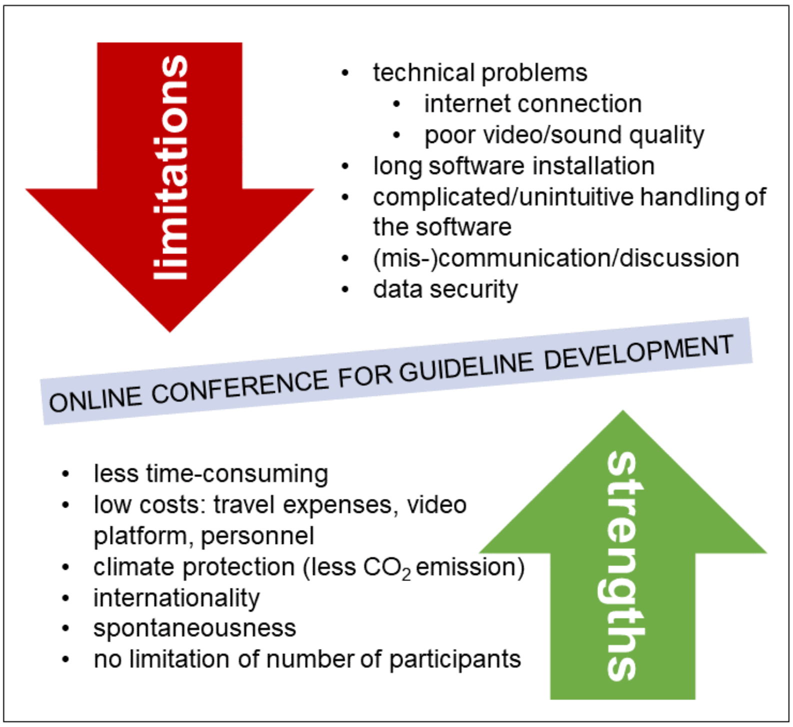
Following the rapid spread of the virus, nearly all countries have introduced temporary restrictions affecting daily life, with “social distancing” as a key intervention for slowing the spread of the virus [5]. Besides these strategies, many companies have introduced or expanded home office–based work, which lead to a tremendous increase in the use of video web conference platforms such as *Skype for Business* (Microsoft Corporation), *Zoom* (Zoom Video Communications), *Google Hangouts* (Google LLC), *Adobe Connect* (Adobe Systems), or *GoToMeeting* (LogMeIn). A web conference or online meeting is a “virtual” meeting organized and conducted over the internet between participants in different locations. In a web conference, the desktop of the meeting moderator replaces the real conference table. All participants can follow what is happening on the moderator’s desktop in a window on their screen, such as holding a presentation or editing a text document. During the web conference the moderator’s role can be flexibly switched between the participants.

Until now, it remains unclear for how long temporary restrictions will be sustained. Despite the pandemic, the development or actualization of medical guidelines especially in the rapidly changing field of oncology needs to be continued to provide up-to-date evidence- and consensus-based recommendations for shared decision making and maintaining the treatment quality for patients. Furthermore, the COVID-19 pandemic itself is likely to affect specific recommendations of guidelines. This explains the urgent need to continue working on current guidelines.

Online conferences may help to replace face-to-face consensus meetings and ensure guideline development in real time while the pandemic is still ongoing. Face-to-face conferences require considerable personnel, financial, and time resources. In contrast, web-based conferences have the potential to reduce time efforts and cut costs, as business travels are unnecessary (Figure 1). The reduction of business travels may also contribute to a reduction in CO_2 emissions, and therefore, online conferences may contribute to climate protection. Besides, collaboration is independent of location and, hence, strengthens international and interdisciplinary guideline work. Web-based conferences are easy to organize and may be scheduled spontaneously. This makes working in teams particularly efficient and may increase the productivity. Additionally, previous surveys showed that online consensus conferences combined with a telephone conference are a feasible and acceptable approach among guideline developers and participants, especially for participants with prior experience with consensus conferences [6,7]. However, technical issues like poor video quality, frozen screens, or choppy sound snippets may hinder guideline developers from realizing online conferences. In a survey among participants for the guideline development of actinic keratosis, the majority reported no technical problems with the participation in the online conference, whereas the possibility for discussion was rated to be possibly inferior to traditional face-to-face conferences [7]. Besides this, further limitations include miscommunication due to poor video quality, sometimes complicated and unintuitive operation of the software, and concerns regarding data security. Another disadvantage includes the lack of social interaction in online meetings as well as the lack of possibilities for networking. A lot of contacts, ideas, and opportunities arise from social interaction before and after meetings, particularly in small group face-to-face communication. In addition, young researchers could miss out on introducing themselves to potential mentors. Furthermore, the communication is only limited to one person at a time, and there is little room for criticism. Besides, difficulties in capturing mood and dynamics of the audience may remain elusive.

In summary, online conferences represent an important and precious tool to advance guideline development, especially during the COVID-19 pandemic but also for the future in the era of digitization. Nevertheless, to counteract possible disadvantages of web conferences, a moderator should be chosen to guide through the meeting and to warrant formal consensus finding processes. Besides, aims of the meeting, technical requirements, and rules for discussion should be defined and communicated in advance. Additionally, software of different providers should be compared to find the most suitable one. Future efforts should be undertaken to establish a framework for successful online conferences for guideline development [8].

Figure 1. Strengths and limitations of online conferences for medical guideline development.



Conflicts of Interest

None declared.

References

1. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
2. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19). Clinical, laboratory and imaging features of COVID-19: a systematic review and meta-analysis. Travel Med Infect Dis 2020;34:101623 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101623] [Medline: 32179124]
3. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Intern Med 2020 Jul 01;180(7):934-943 [FREE Full text] [doi: 10.1001/jamainternmed.2020.0994] [Medline: 32167524]
4. Mahase E. Covid-19: WHO declares pandemic because of "alarming levels" of spread, severity, and inaction. BMJ 2020 Mar 12;368:m1036. [doi: 10.1136/bmj.m1036] [Medline: 32165426]
5. Kaspar K. Motivations for social distancing and app use as complementary measures to combat the COVID-19 pandemic: quantitative survey study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
6. Nast A, Rosumeck S, Sporbeck B, Rzany B. [Using new media for online consensus conferences and open external review of guidelines - results of two pilot studies]. Z Evid Fortbild Qual Gesundhwes 2012;106(4):295-301. [doi: 10.1016/j.zefq.2011.09.023] [Medline: 22749078]
7. Werner RN, Jacobs A, Rosumeck S, Nast A. Online consensus conferences for clinical guidelines development - a survey among participants from the International Guidelines for the Treatment of Actinic Keratosis. J Eval Clin Pract 2014 Dec;20(6):853-856. [doi: 10.1111/jep.12159] [Medline: 24799003]
8. Reshef O, Aharonovich I, Armani AM, Gigan S, Grange R, Kats MA, et al. How to organize an online conference. Nat Rev Mater 2020 Mar 18:1-4 [FREE Full text] [doi: 10.1038/s41578-020-0194-0] [Medline: 32218994]

Edited by G Eysenbach, R Kukafka; submitted 11.05.20; peer-reviewed by M Kirchberger, B Cao, H Rjoub; comments to author 31.08.20; revised version received 07.09.20; accepted 02.10.20; published 29.12.20.

Please cite as:
Steeb T, Follmann M, Hagen RM, Berking C, Heppt MV
Implications of the COVID-19 Pandemic for the Development and Update of Clinical Practice Guidelines: Viewpoint
J Med Internet Res 2020;22(12):e20064
URL: http://www.jmir.org/2020/12/e20064/
doi:10.2196/20064
PMID:33347419

Tutorial

Developing Virtual Reality Trauma Training Experiences Using 360-Degree Video: Tutorial

Devika Patel[1], MS; Jessica Hawkins[2], MSED; Lara Zena Chehab[1], MPH; Patrick Martin-Tuite[3], MS; Joshua Feler[4], MD, MS; Amy Tan[5], BA; Benjamin S Alpers[1], MS; Sophia Pink[6], MS; Jerome Wang[7], BA; Jonathan Freise[1], MD, MS; Phillip Kim[1], MD, MBA; Christopher Peabody[8], MD, MPH; John Bowditch[9], MA; Eric R Williams[9], MFA; Amanda Sammann[1], MD, MPH

[1]Department of Surgery, University of California, San Francisco, San Francisco, CA, United States
[2]School of Medicine, Stanford University, Stanford, CA, United States
[3]School of Medicine, University of California, San Francisco, San Francisco, CA, United States
[4]Department of Neurosurgery, Brown University, Providence, RI, United States
[5]School of Medicine, University of California, Davis, Sacramento, CA, United States
[6]School of Engineering, Stanford University, Stanford, CA, United States
[7]School of Public Health, University of California, Berkeley, Berkeley, CA, United States
[8]Department of Emergency Medicine, University of California, San Francisco, San Francisco, CA, United States
[9]J. Warren McClure School of Emerging Communication Technologies, Scripps College of Communication, Ohio University, Athens, OH, United States

Corresponding Author:
Devika Patel, MS
Department of Surgery
University of California, San Francisco
1001 Potrero Ave
San Francisco, CA, 94110
United States
Phone: 1 628 206 3764
Email: devika.patel@ucsf.edu

Abstract

Historically, medical trainees were educated in the hospital on real patients. Over the last decade, there has been a shift to practicing skills through simulations with mannequins or patient actors. Virtual reality (VR), and in particular, the use of 360-degree video and audio (cineVR), is the next-generation advancement in medical simulation that has novel applications to augment clinical skill practice, empathy building, and team training. In this paper, we describe methods to design and develop a cineVR medical education curriculum for trauma care training using real patient care scenarios at an urban, safety-net hospital and Level 1 trauma center. The purpose of this publication is to detail the process of finding a cineVR production partner; choosing the camera perspectives; maintaining patient, provider, and staff privacy; ensuring data security; executing the cineVR production process; and building the curriculum.

(J Med Internet Res 2020;22(12):e22420) doi:10.2196/22420

KEYWORDS

virtual reality; cineVR; 360-degree video; trauma training; medical education

Introduction

Virtual Reality in Medical Education

Emergency and inpatient medical training has historically occurred in an apprentice style of teaching. The famous adage “see one, do one, teach one” describes the process where trainees would observe a procedure once, perform that procedure themselves, and subsequently teach others how to perform it. However, this learning process traditionally takes place in the context of real patient care, which inevitably results in increased medical errors as the learner practices their skills [1].

Fortunately, over the last 2 decades, medical education has embraced the use of simulated environments as a way to improve the competency of trainees while decreasing medical errors. This not only increases patient safety but also improves the learner's clinical skills by allowing for repetitive practice of the intended skills coupled with specific and informative feedback that results in better skills performance [1-3].

Simulation-based medical education uses actors, mannequins, or trainers to create an artificial representation of a real-world process to facilitate experiential learning [4]. Simulation has been shown to improve teamwork and communication across high-risk industries and a variety of medical disciplines including anesthesiology, surgery, obstetrics, emergency medicine, pediatrics, and critical care [5,6]. Although medical simulation is an effective and well-established teaching tool, most methods depend on expensive models, require specific equipment and manpower, and are commonly considered to be low-fidelity and artificial to allow suspension of disbelief and immersion in the simulation [6-10].

Virtual reality (VR)–based medical education offers a potential solution to these traditional simulation limitations. VR allows learners to immerse themselves in an authentic and accessible clinical scenario using 360-degree video VR and audio technology (cineVR), without compromising patient safety [11,12]. CineVR is a subset of VR that depicts a real (versus computer-rendered) environment where the viewer can look around the room and see in all directions but cannot interact directly with the scene. Unlike traditional simulation-based education, cineVR allows for flexibility, independent learning, and unlimited number of opportunities to practice and hone teamwork, communication, and patient management skills. While in cineVR, the learner is not interacting with a virtual environment, but with the authentic, real-world environment of the 360-degree video. CineVR allows learners to gain familiarity with the environment before being there in real life, and a qualitative study has shown that the realism associated with cineVR can elicit emotional responses [13].

A growing number of academic medical centers are using VR to augment their training curricula [14,15]. VR technology is being used widely in surgical education. Several randomized controlled trials have found that practicing surgical skills in VR improved technical performance in the operating room [16,17]. CineVR has been used in a variety of settings, including naloxone administration training, where it was shown to improve knowledge and perceptions about opioids and use of naloxone, and in diabetes education, where it was shown to improve cultural self-efficacy and diabetes attitudes among health care providers and administrators [13,18]. These technologies are also being applied with companies such as Embodied Labs, exploring educational topics such as patient communication. An early study suggests that participants that watched one of the Embodied Labs simulations had positive learning experiences, demonstrated empathy, and reported an increase in their perceptions of using 360-degree VR as a learning tool [19].

Virtual Reality in Trauma Care

Medical and surgical trainees have limited opportunities to safely practice trauma response skills without compromising patient safety. Trauma care is defined as care for critically injured patients, where patient care scenarios can include major traumatic injuries such as motor vehicle collisions or gunshot wounds [20]. High-level trauma resuscitations require coordination across multidisciplinary care teams in an unpredictable and chaotic environment. For most medical trainees, trauma care training occurs on the job, with the exception of a 2-day Advanced Trauma Life Support (ATLS) training course that is renewed every 2 years and includes 1 day of simulation training with standard actors. The ATLS course focuses on the process and necessary procedural skills to evaluate and care for an injured trauma patient. While surgical residency programs sometimes implement simulation-based trauma care practice, these are limited by cost and duty-hour regulations and not guided by standardized best practices or strategies [21]. The majority of trauma instruction is therefore variable, with trainees learning through experiential learning as an observer or participant in live traumas.

This method of trauma training is suboptimal for 3 reasons. First, it leads to a cohort of trainees with variable experience and facility in managing trauma patients [21]. Second, it puts learners in a highly stimulating and stressful environment without experience or stress management techniques, which can decrease clinical decision-making capacity [22,23]. Third, learners often do not receive adequate training in leading, managing, or coordinating the other medical providers in the room [24].

Project Goal

The goal of this project was to develop a trauma training curriculum, to be used as standalone or part of standard ATLS training, for residents by using 360-degree videos of real trauma care scenarios, to address the limitations of traditional trauma care training. The curriculum focuses on skills such as teamwork, the management of a trauma patient, understanding and empathizing with the roles and priorities of different team members, and anticipating next steps in care planning. This article specifically discusses the process and challenges related to (1) finding a production partner, (2) selecting recording perspectives, (3) equipment selection, (4) setting up and filming the 360-degree videos, (5) file management and video editing, (6) data protection and privacy, and (7) curriculum development.

In this article, we describe our process for planning, filming, and producing trauma care training videos. We hope that, in doing so, we demonstrate the feasibility for a small group to explore VR simulation and encourage other groups to undertake more VR filming projects. To that end, we highlight both challenges and lessons learned, as well as our recommendations for future work, so that this paper might be used as an open reference tool for future projects.

Methods

We filmed 20 live and simulated high-level trauma resuscitations in one trauma bay at the Zuckerberg San Francisco General hospital over 6 months. This project was a collaboration between the departments of trauma surgery, anesthesia, emergency medicine, nursing, and hospital facilities, and filming and editing were performed by colleagues at the Game Research and Immersive Design (GRID) Lab at Ohio University. Funding for this effort was obtained through the San Francisco General Hospital Foundation. Human subjects protection was granted through the Institutional Review Board (IRB) of University of California, San Francisco. The resuscitations were filmed from 6 different perspectives to capture the experiences of 6 different types of providers. The footage from 8 of these trauma scenarios was used to produce cineVR videos, 4 of which were further edited to include a trauma curriculum for trainees. In this section, we describe our methodology for making these videos, from planning to execution.

Choosing a Production Partner

Most health systems do not have the experience or skills to record, edit, and produce 360-degree video and audio simulations. Therefore, video production necessitated collaboration with an outside partner. This required additional funding, which our team was able to procure through grants.

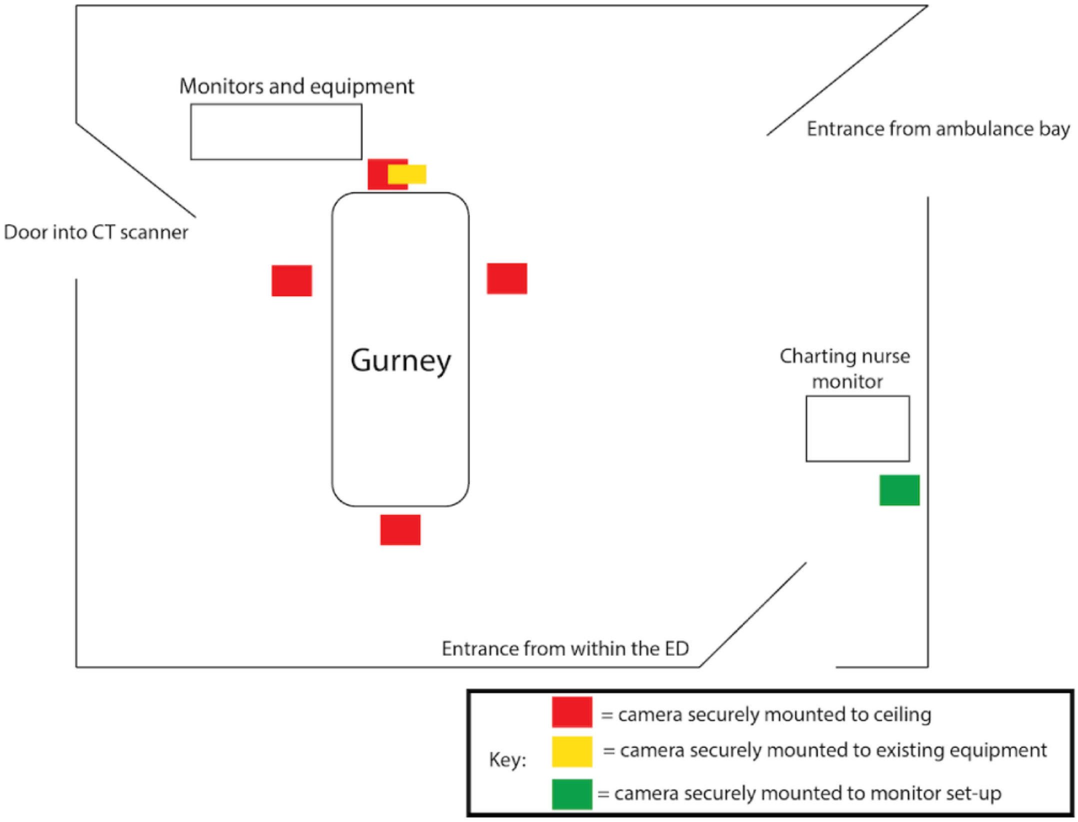
Due to the highly sensitive nature of storing real patient scenarios, a significant challenge required identifying a partner institution with a Health Insurance Portability and Accountability Act (HIPAA)–compliant data security system. To find a partner, we conducted an online search, asked experts in the VR field, and reached out to both research centers and production companies to understand their capabilities, data security infrastructure, cost, health care experience, and timeline. A search of local companies revealed that commercial VR production entities rarely had HIPAA-compliant data editing and storage systems. After an extensive search, we identified the GRID Lab at Ohio University. The GRID Lab, created in 2005, focuses on the research and development of Virtual, Augmented, and Mixed reality experiences; serious and educational games; simulations; computer animation; and motion capture [25]. As an academic institution, they possessed extensive experience in working in the health care space developing cineVR technology. Thus, their institution had standardized infrastructure for being HIPAA-compliant.

Selecting the Recording Perspective

Our project aimed to increase teamwork and empathy by allowing learners to experience the perspectives of their colleagues through cineVR. Therefore, an essential component to early stage planning was selecting the viewer perspectives that might allow for this experience. To allow viewers to experience their colleagues' perspective, we identified the key personnel to capture in the cineVR as (1) team "leader" (usually a trauma surgery resident or emergency medicine resident), (2) nurse, (3) airway provider (usually an anesthesia resident), (4) attending physician, (5) charting nurse, and (6) patient. We intended to capture the traumas from each of these perspectives, generating 6 videos for every trauma event.

In the general practice of trauma care, specific providers often stand at certain sides of the patient. For instance, airway management will primarily be at the head of the bed. Therefore, we chose our camera locations based on the providers we had identified and mounted cameras not only to the left of the bed and the head of the bed, but the right of bed (nurse perspective), foot of bed (attending physician perspective), periphery of the room (charting nurse perspective), and the patient gurney (patient perspective; Figure 1).

Figure 1. Schematic of equipment set-up in the selected trauma resuscitation room. CT: computed tomography; ED: emergency department.



Equipment Selection Considerations

There are many commercially available cameras with the ability to film in 360 degrees. We considered the following when evaluating potential equipment.

Image Quality

Quality varies greatly between cameras and at different price points. To make our videos as immersive as possible, we chose cameras that record video in 4k (a resolution of 3840 x 2160 pixels, compared to standard HD video of 1920 x 1080 pixels) to ensure high-fidelity output.

Auto-Stitching

360-degree videos are created by "stitching" together 2 or more separate video files. Manually stitching these files together is resource-intensive and time-intensive; therefore, we chose a camera with an "auto-stitch" feature, which allows for simple and streamlined video processing.

Equipment Size

Cameras are not the only equipment needed for filming; clamps, mounts, and booms are also a standard part of the unit. In an already-cramped resuscitation room, extra equipment can easily disrupt workflow. As we needed 6 equipment units in the room at once, we selected cameras that were as small as possible.

Ease of Use

To expedite filming, we preferred equipment that was intuitive to operate and required minimal additional training.

Equipment Selection

After considering these 4 features, we selected Yi Technology 360 VR cameras (Yi Technology, Shanghai, China). The cameras produce high-quality images, have an "auto stitch" ability, are small and lightweight, and are easy to use. Additionally, the cameras allow for remote monitoring by streaming to a smartphone. We paired each camera to a Samsung Galaxy (Samsung C&T Corporation, Songpa-gu, Seoul, South Korea) smartphone, which functioned as a monitor. These smartphones were kept offline for data security.

We similarly selected Zoom H2N (Zoom Corporation, Chiyoda City, Tokyo, Japan) microphones because of their high-quality recording (96 kHz sample-rate/24-Bit resolution in 360 degrees), small size, and intuitive design. These microphones recorded audio in an "ambisonic," or 360-degree pattern, as opposed to "stereo," which records in only 2 directions. When an ambisonic audio recording is stitched to a 360-degree video, the audio and video replicate a real-world experience. The audio is able to rotate in relationship to the visual object that is creating the sound. For example, if someone on one's right is talking in the video, one would hear a voice coming from their right-hand side.

To ensure the audio and video matched, the audio recorders were mounted in the same location as the cameras. All video and audio were downloaded to a laptop purchased specifically for this project. We chose an MSI laptop (Micro-Star International Co. Ltd., New Taipei City, Taiwan), a popular gaming laptop brand, for its high-speed performance and reliability. The MSI laptop has an Intel Core i7-10750H 2.60 Ghz processor, 8GB of RAM, internal 256 GB M.2 2280 Plce NVMe SSD storage, the GeGorce GTX 1650, and graphics card.

Additional detailed descriptions of our decision-making process on choosing the cameras and microphones are in Multimedia Appendix 1.

Setting Up and Filming the 360-Degree Videos

We selected a single resuscitation room for filming. This section describes our considerations and approach to positioning and mounting the equipment and to testing the room set-up to ensure it did not interfere with provider workflow.

Room Set-Up

To set up the room, we had to consider camera positioning, ceiling hanging equipment, and mounting to existing equipment.

It was necessary to ensure that the equipment would not interfere with patient care. In selecting camera position, we primarily considered provider workflow, selecting locations that would not disrupt view of or access to the patient. However, we also considered the potential for the camera to obstruct the path of a large portable machine, such as an x-ray machine. As an additional safety measure, we planned to mount our equipment such that any camera could be rapidly pushed out of the way in the case of an unexpected emergency.

Our production partner visited several months before filming to select tentative camera positions and to prototype equipment mounting systems. We selected 2 methods to mount equipment, depending on the camera location: hanging equipment from the ceiling or mounting on existing equipment in the room. To ensure compliance with hospital policy, the director of facilities met with our team to discuss rules and weight restrictions for hanging camera equipment.

Cameras capturing the "head of bed," "right of bed," and "left of bed" perspectives were hung from the ceiling. We obtained approval from the hospital facilities department for our ceiling-mounted cameras. We mounted the camera using an Impact brand clamp (Gradus Group LLC, Secaucus, NJ), which was attached to boom poles and hung from ceiling hatches using wires (Figure 2). We weighed this setup to ensure it did not surpass weight restrictions. The entire ceiling mount was then covered with white gaffer's tape in an attempt to make it less distracting to providers. All clamps and all extension arms had an additional safety tether for added security.

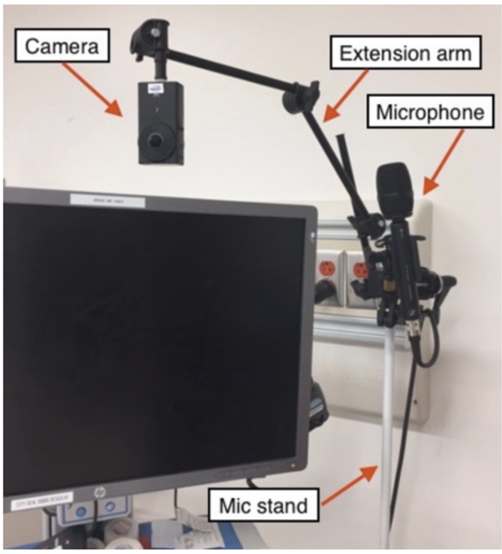
Figure 2. Ceiling camera configuration.

The additional camera and microphone units were mounted to existing equipment in the resuscitation room.

A camera was placed at the foot of the bed to capture the perspective of an attending physician standing at the foot of the bed; we mounted a camera to an overhead lamp just over the attending physician's head. To do this, we used a clamp to attach the camera to a 4-inch long, 0.5-inch diameter metal rod; then, we used white gaffer's tape to attach this rod to the lamp. A safety wire was added to attach the camera to the lamp arm.

To mount the camera at the charting nurse station, we used a Sennheiser AMBEO microphone unit (Sennheiser electronic GmbH & Co, Hanover, Germany) on a weighted base (7 pounds), which we placed on the floor next to the charting nurse station. We affixed the camera to an extension arm, which we then directed over the computer monitor to approximate the point of view of the nurse (Figure 3).

Figure 3. From the charting nurse perspective, the equipment configuration using a mic stand and extension arm.



The camera angle from the patient gurney attempted to capture the view of the patient looking up at the trauma care team. We attached the camera and a complete Sennheiser AMBEO microphone unit to a Manfrotto camera extension arm (Manfrotto, Ramsey, NJ), which was affixed to the underside of the gurney using an Impact Super Clamp.

Testing the Filming Set-Up

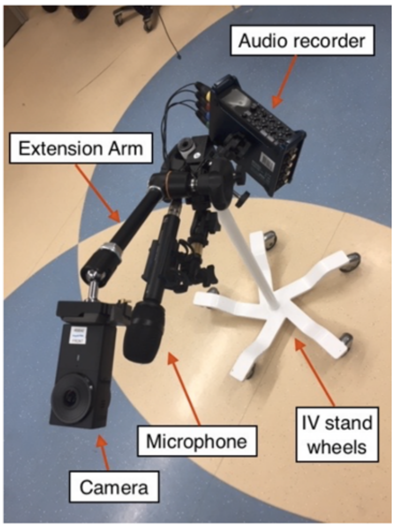
Prior to recording, with the approval of the emergency department leadership, we closed the previously selected resuscitation room for 2 hours to run trauma simulations with emergency medicine, surgery, anesthesia, nursing, and radiology teams to obtain feedback on the camera position. This feedback was used to revise positioning of the equipment. We also reviewed sample video from each camera position to ensure it captured the appropriate perspective while remaining nonobstructive to patient care.

Two unforeseen issues that were addressed during this testing were provider height and obstruction of the airway-focused provider's workflow.

Regarding provider height, cameras were attached between a 5'11'' to 6'2'' height when possible (to approximate the view from a someone of that height viewing the patient). Providers at or above this height occasionally found ceiling-mounted cameras interfered with their space. To address this, taller providers were briefed before they entered the trauma bay. Overall, most providers and staff acclimated quickly.

Regarding airway-focused providers, capturing the anesthesiologist perspective was difficult because the provider contended both with the “head of bed” camera from the ceiling and the “patient perspective” camera attached to the gurney. To mitigate the equipment obstructing the anesthesiologist’s workflows, the production team mounted the camera and microphone unit to a wheeled IV stand, allowing the provider to push the patient perspective camera out of the way when necessary (Figure 4).

Figure 4. From the patient perspective, the equipment configuration using an IV stand on wheels.



Filming Process

We filmed over 2 periods in a single trauma resuscitation room, first for 10 hours per day for 4 days and subsequently for 24 hours a day for 6 days. The 2 periods of filming were chosen to maximize the potential number of potential resuscitation events. Due to the unpredictable nature of trauma care, we were unable to capture the desired library of trauma cases during this initial filming period. We therefore planned a second, longer film period. During the second period, we increased filming efforts from 10 hours per day to 24 hours per day. While our goal was to capture real patient-care scenarios, we had anticipated not being able to capture our proposed set of videos in real life and recorded simulated traumas with patient actors using a real trauma team in the resuscitation room before each of the 2 live filming periods began. By filming simulations prior to capturing the live resuscitations, we were able to practice what real filming would entail, allow for the production team to have a dry run in the trauma bay to learn the workflows, and allow for any last-minute issues to be resolved. It was necessary to ensure that filming was not going to compromise real trauma care.

The team was assigned to work on production in 10- to 12-hour shifts. Each shift was composed of a 6-person team that included 3 production staff, 1 supervisor, and 2 research assistants. We created a control center near the resuscitation room where the 3 production staff monitored the equipment during recording: 2 focused on video and 1 focused audio (Figure 5). The control center also housed 1 extra set of equipment, kept fully charged and with memory cards available, 1 cell phone to monitor each camera, and a data log for all files and notes for each case. When the team was alerted of a trauma, they would enter the room and quickly turn on all 6 cameras and microphones.

Figure 5. CineVR Control Station, where all equipment was placed during filming.

Equipment Redundancy

After each trauma, cameras and microphones needed to be cleaned, batteries and memory cards replaced, and data uploaded to the secure server. In order to be ready for filming whenever a trauma entered the room, our team employed a strategy of equipment redundancy. Six cameras with corresponding microphones were positioned in the trauma bay, another full set was on stand-by in the control center, and 3 extra camera and microphone units were available as backup. Following each trauma event, the production team swapped out the cameras and microphones in the room to ensure new equipment was ready to use in the case of another trauma.

File Management and Video Editing

Each filmed trauma event generated multiple files. Once videos were appropriately named and catalogued on an encrypted laptop, they were transferred to a secure, HIPAA-compliant, cloud-based storage system. In this section, we detail our file management system and review our brief video-editing process.

File Management

As described in earlier sections, the Yi camera automatically stitches together 2 video inputs into a single 360-degree video file. The maximum length of this video file is 4 minutes. After 4 minutes, the camera continues to record, but it generates a new file. Therefore, during an 18-minute recording, the camera creates 5 files: 4 files with 4-minute duration and 1 file with 2-minute duration. The Zoom microphone records audio as a single 360-degree audio file with no maximum length. Therefore, each camera and microphone unit recording an 18-minute trauma produced 6 total files: 5 video files and 1 audio file. For every trauma, we recorded from 6 camera and microphone positions. Hence, an 18-minute trauma would produce 36 separate files, 6 from each camera and microphone unit. To organize these files, we created a systematic labelling system, with patient number, recording device, location in room, and file sequence (A, B, C, D, etc).

Video Editing

Our goal was to replicate the experience of being in a real trauma room; therefore, we did not alter the video content to preserve the authenticity of the experience. We used Adobe Premiere Pro (Adobe Inc, San Jose, CA) to combine the disparate video files and any other video editing. For each video, the individual camera files were organized into the correct chronological sequence, then synchronized with the microphone audio file. Visible entryways were blurred to ensure unconsented passersby would not be included in the final product. A title card and countdown were added to the beginning of the video. The files were then transferred to Oculus Go (Oculus, Menlo Park, CA) VR headsets for testing, chosen for their portability and low cost.

The production partner and research team must align and work together to ensure that file management is well-understood and data security guidelines are followed by both teams.

Data Protection and Privacy

It was essential to ensure patient privacy because patient care was being captured on camera to create the curriculum. IRB approval was obtained through the University of California, San Francisco. The IRB addressed the following issues: consent, data collection, storage of data, sharing of data, editing of data, and study procedures.

Obtaining Consent

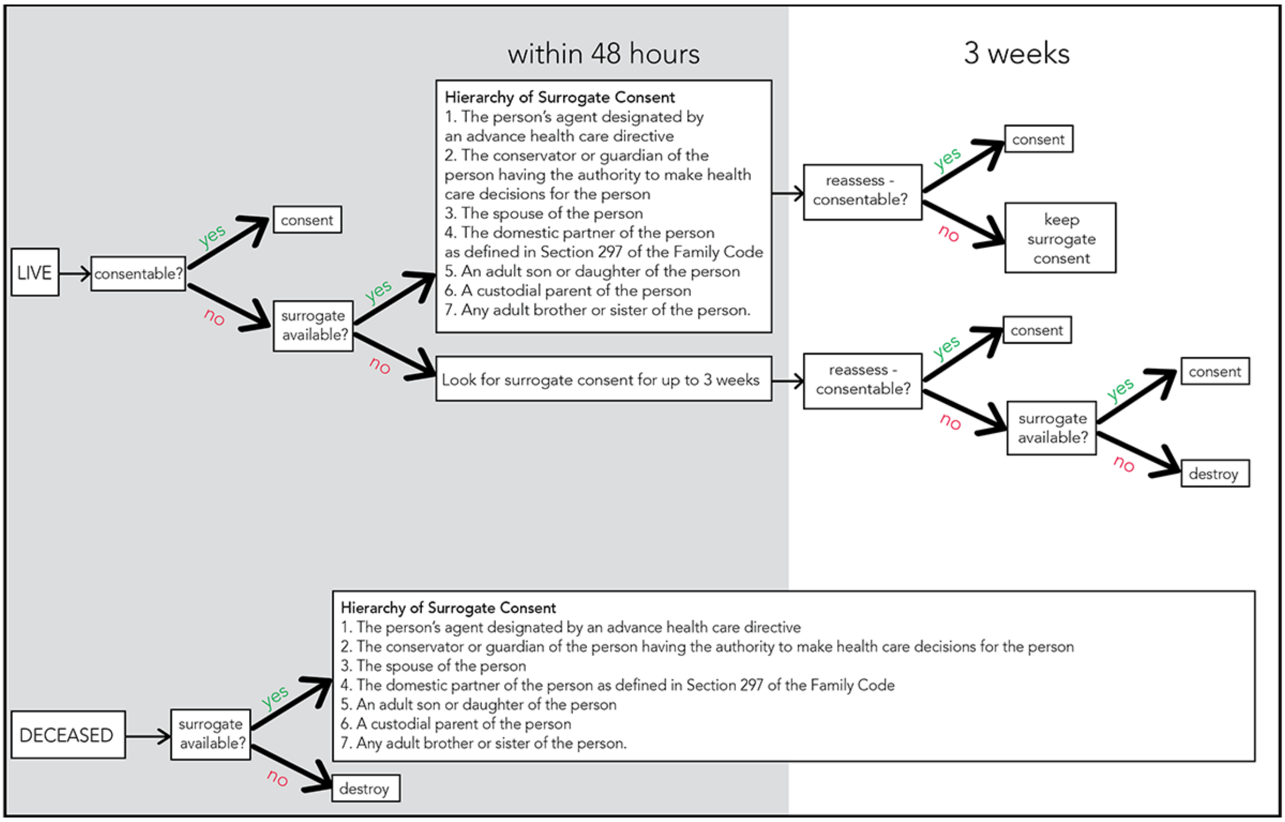
We needed to obtain consent from everyone who would be captured on film. This included patients, providers, other hospital staff, and emergency medical services (EMS) personnel.

Developing a process to consent all providers responding to the trauma was a challenge and demanded proactive and retroactive effort by research staff. We sent emails describing the study to providers prior to their shifts. We also worked with the administrator on duty and nursing managers to organize in-person meetings during provider shifts about the filming project. During these meetings, study coordinators described the study, answered questions, and obtained signed consent. Providers and staff who did not provide consent were reassigned to a different resuscitation room for the shift when filming was occurring.

We were unable to get consent from some providers prior to filming, such as EMS personnel. To obtain retroactive consent in these cases, we stationed 1 research assistant at each of the room's 2 entrances to identify unconsented providers and obtain informed consent.

The time-sensitive nature of trauma care prevented our team from consenting the patient prior to filming. When possible, we requested consent from the patient after filming. If the patient was unable to provide consent due to severe disability, we attempted to obtain consent from a surrogate. In these instances, we referred to our institution's hierarchy of surrogate consent. In other cases, we would reassess patient status and seek to obtain consent directly from the patient up to 3 weeks after filming, described further in Figure 6.

Figure 6. Consenting structure for patients, detailing the hierarchy of surrogate consent.



Hierarchy of Surrogate Consent
1. The person's agent designated by an advance health care directive
2. The conservator or guardian of the person having the authority to make health care decisions for the person
3. The spouse of the person
4. The domestic partner of the person as defined in Section 297 of the Family Code
5. An adult son or daughter of the person
6. A custodial parent of the person
7. Any adult brother or sister of the person.

Hierarchy of Surrogate Consent
1. The person's agent designated by an advance health care directive
2. The conservator or guardian of the person having the authority to make health care decisions for the person
3. The spouse of the person
4. The domestic partner of the person as defined in Section 297 of the Family Code
5. An adult son or daughter of the person
6. A custodial parent of the person
7. Any adult brother or sister of the person.

Secure Data Collection and Storage

We consulted with data and security officers in the information technology and compliance departments at our institution to develop a secure data storage and transfer process.

After each filmed trauma event, captured media files were uploaded from the devices' SD memory cards to a laptop and backed up to a hard drive. Files were then deleted from the SD cards. At the end of each day, files were uploaded from the encrypted laptop to a HIPAA-compliant, cloud storage system. Once stored in the cloud, all files were deleted from the hardware.

Video editing was performed at the GRID Lab at Ohio University. The videos were downloaded from the cloud storage onto an encrypted hard drive, which was used to transfer the files over a secure network to a HIPAA-compliant computer in a private editing room reserved specifically for this project. The videos were edited on these secure computers, transferred back to the cloud storage system, and then deleted from all hardware.

Access to the cloud storage system was limited to 3 personnel at the GRID lab, and a maximum of 2 videos were downloaded at a time. All data repositories, including laptops, hard drives, and cloud-based storage, were encrypted in accordance with institutional protocol.

To ensure data security and privacy, work closely with your institutions' IRB office, legal team, and data and security officers in the information technology and compliance departments of your institution to craft an IRB and appropriate partnership agreement that complies with all data security measures.

Curriculum Development

The videos were reviewed by the research team, and 8 videos were selected to be further edited based on the visual quality and educational potential or the applicability to various trauma care providers of the video. A multidisciplinary team was assembled that included the research team, a trauma surgeon, an emergency medicine physician, an anesthesiologist, and a Vice Dean of Medical Education to develop a curriculum to overlay on the videos.

Using Cognitive Load Theory to Inform the Curriculum

Principles of Cognitive Load Theory (CLT) were used to build the curriculum. CLT posits that learning happens best when instructional materials direct cognitive resources towards activities relevant to learning [26]. CLT describes 3 types of cognitive load: intrinsic load (essential to the learning task), extraneous load (nonessential to the learning task and often caused by poor design), and germane load (work put in to enhance schema formation). This framework was used to augment the 360-degree videos with visual and auditory overlays to train learners to focus on relevant inputs and to ignore the irrelevant inputs. Relevant and irrelevant inputs were defined through learner and expert interviews. To identify the irrelevant stimuli that contribute most significantly to the extraneous load, 28 junior residents from surgery, emergency medicine, and anesthesia were asked to view 2 trauma resuscitation videos in a VR headset. Each resident was given 4 prompts to think about and reflect on while watching the video, known as a "talk aloud." These prompts asked learners to explain what they are seeing and thinking while watching the VR video, what they believe to be their priorities in managing care for the patient, what they find to be essential stimuli, and what they find to be distracting stimuli. After each VR viewing session, residents participated in a semistructured interview.

The 8 videos were also viewed by an expert physician within each discipline in surgery, emergency medicine, and anesthesia. These expert physicians also participated in "talk alouds" and reflected on aspects of the videos they believed learners might find challenging, stressful, distracting, and overwhelming. They also identified key learning objectives for junior residents in their respective specialties.

The residents and attending physician responses from their "talk alouds" and semistructured interviews were analyzed by the research team to identify thematic sources of cognitive load. We identified 5 sources: (1) acclimating to the chaos of the room and understanding the roles and priorities of the trauma team, (2) understanding and applying the protocol for evaluating and caring for a trauma patient per the ATLS guidelines, (3) anticipating next steps in care to manage the patient, (4) understanding clinical content, and (5) orienting to VR. Based on their content, 2 videos were then assigned to each of the first 4 categories. One video was created to orient the learner to experiencing the capabilities of cineVR.

Between 3 and 8 teaching points were created for each video and synchronized to the appropriate time in the video. The editing team created graphics using Adobe Photoshop software. The graphics were then animated using Adobe After Effects software before importing the animated graphics into the Adobe Premiere Pro editing software. For 2 of the videos, a script was written, and audio was recorded to be inserted as "voiceover." The augmented videos were reviewed by the attending physicians, and further edits were made to improve timing and content.

Lessons Learned

After going through the processes of preparation, set-up, filming, editing, and curriculum development in creating the trauma-training cineVR videos, our team has gathered a set of recommendations to guide other medical institutions interested in creating such a curriculum. Overall, our team learned the importance of working collaboratively and closely with both production partners and institutional partners, and testing equipment and filming scenarios beforehand allow the entire team and all stakeholders to be prepared and ready for filming day. Thinking through the logistics and workflows for each process, along with gaining appropriate approvals, can help streamline your filming and production and yield valuable cineVR videos. In Textbox 1, we summarize our recommendations for readers looking to embark on a similar process.

Textbox 1. Planning recommendations for each of the steps to create a 360-degree video and audio (cineVR) curriculum.

Finding a production partner

- Conduct a broad search to identify a partner, including an online search, a literature search, and reaching out to virtual reality experts. Depending on your specific needs, your partner might be an industry partner or an academic center.
- Prioritize production partners with experience in health care.
- Ensure that they have the necessary technical infrastructure to maintain data security and Health Insurance Portability and Accountability Act (HIPAA) compliance.
- Seek partners who are open to collaboration.

Selecting recording perspectives

- Collaborate with a diverse cohort of physicians and staff to identify the perspectives that must be captured.

Equipment selection

- Evaluate the 360-degree video and audio recording technologies to optimize for audio and image quality, auto-stitching capability, size, and weight, as well as ease of use. Additional features to consider include external or internal battery and cost.

Setting up and filming the 360-degree videos

- Iteratively test different camera configurations in the room with providers and staff to ensure that the equipment does not get in the way of patient care. Run simulations of the patient care scenario to identify whether cameras have the potential to obstruct care or capture suboptimal video.
- Partner with the facilities team at your institution, as they can provide additional support, equipment, and advice to ensure that the equipment is tethered safely and within institutional guidelines.
- Avoid technical complications by having 2 sets of all filming equipment.

File management and video editing

- Create a standard file management protocol, followed by both the research and filming team.
- Appoint a data security lead that can communicate with the filming team and ensure file organization throughout the process.
- Blur any parts of the video that involve information or an individual that was not consented for (eg, an individual, not part of the filming, who walks past an open door).

Data protection and privacy

- Consult your institution’s data security, compliance, information technology, and legal teams early in the planning process to develop a data security and monitoring plan that adheres to all necessary guidelines.
- Develop a patient consent plan that tracks each patient to every part of their hospital stay.
- Co-develop a surrogate consent protocol with your institution’s institutional review board.
- Obtain consent from providers, staff, and other participants before filming, when possible. Attend change of shift and staff meetings to present the project, answer questions, and obtain consent.
- To capture any remaining providers, staff, police, emergency medical services, or any other individuals who might be filmed, station research staff outside each door of the trauma bay to ensure that all participants are consented.

Curriculum development

- Collaborate with educators, learners, and other stakeholders to iteratively design the curriculum.

Discussion

This study demonstrates the feasibility of developing a cineVR-based training curriculum using real patient-care scenarios at an academic medical center. We illustrate important lessons learned from our experience in the hope other institutions may refer to this publication when planning similar projects. Important considerations include finding a production partner, choosing individual camera perspectives, setting up and filming the 360-degree videos, video editing and file management, data protection and privacy, and building the curriculum.

While this process requires significant time investment by medical and research personnel and extensive coordination across multiple stakeholders, these videos can be used for a variety of curricular needs. In the long-term, such video libraries, capturing a breadth of typical cases in trauma care, can be shared or exported to other trauma centers and teaching hospitals nationally.

The use of cineVR technology for medical simulation training is unique and promising because it addresses a major critique of traditional medical simulation: the lack of fidelity to real-life situations [10]. Currently, no simulation realistically depicts all of the physiological, mental, and behavioral components of patient care [27]. Poorly designed simulations can neglect important components of patient scenarios, suggesting that those missing components are unimportant or unintentionally encouraging shortcuts [28].

In studying simulation training in other disciplines, it has been shown that psychological fidelity is more important than physical fidelity [10]. The videos described in this article illustrate 2 separate but related psychological concepts in the virtual reality literature: "presence" and "PREality." "Presence" is a person's subjective sensation of being in a (usually virtual) scene, enabling them to interact with and feel connected to the world outside their physical bodies via technology [29]. "PREality" is designed to prepare the viewer for an anticipated experience by placing them in the actual environment where they intend to work [30,31]. It forces a sense of déjà vu on the viewer, allowing them to connect more closely to a specific environment or activity. This is a different psychological phenomenon than the idea of presence. Although 360-degree videos do not have to be location-specific to create a sense of presence, they do need to be location-specific to create a sense of "PREality." The methods to produce cineVR videos described here use both presence and PREality to establish a high level of psychological fidelity.

Until now, the extent to which cineVR has been used in medical education has been limited, with videos focused primarily in an operating room setting, using real or simulated patients [11,32]. Educational research on the potential of cineVR to augment trauma response training is extremely limited; to our knowledge, only one other study has been published [33]. In this study, nursing students viewed 360-degree video of trauma teams in Sweden using a desktop computer, and the analysis found that the tool can be a useful addition to existing methods in nursing education.

Limitations

The video experiences produced at our institution may not be as applicable to learners at other institutions because of differing physical environments. Given the difference in environments, learners from other institutions will retain the sense of presence but will not obtain a sense of "PREality" when training through our curriculum. To combine presence and "PREality" with learning, institutions should develop their own training curriculum that is based in their particular environments.

Next Steps

Our next steps are to test the feasibility, acceptability, and implementation of the curriculum described here in a proof-of-principle study among medical trainees. The specific details of the curriculum will be outlined in future publications. CineVR is a novel educational tool, and additional research can address several unanswered questions. For instance, there is limited research suggesting how to test the effectiveness of such a curriculum, how this technology can be used in team training, and the feasibility of incorporating additional capabilities such as gaze-tracking.

Conclusion

It is feasible to develop a VR-based training curriculum using video from real patient-care settings. To do so, it is integral to partner with experienced organizations, plan and test filming processes, ensure robust data and physical security of all video and audio components, coordinate multistakeholder efforts, and be agile in unpredictable and uncharted environments. Although further research is required to understand the feasibility and effectiveness of such a curriculum, we hope the learning we gained from this process can guide other academic teaching hospitals and trauma centers who aim to develop high-fidelity training for their medical education programs.

Acknowledgments

We would like to acknowledge and deeply thank everyone involved in this project, including the patients, providers, and staff who participated in the filming and supported this study. We would also like to thank Pamela Derish for her support in editing the manuscript. AS conceived the study, and the study was designed by AS and CP. The team was led by AS and DP, with support from LC, PM-T, JF, SP, JW, JF, and PK. Technical specifications and equipment were managed by JB and EW. The paper was drafted by DP, JF, EW, JH, and AS. All authors reviewed and revised the manuscript.

Conflicts of Interest

AS, LC, and DP are human-centered design consultants for The Empathy Studio, LLC. All other authors declare they have no competing interests.

Multimedia Appendix 1
Equipment specifications for the cameras and microphones.
[DOCX File , 24 KB - jmir_v22i12e22420_app1.docx]

References

1. Al-Elq AH. Simulation-based medical teaching and learning. J Family Community Med 2010 Jan;17(1):35-40 [FREE Full text] [doi: 10.4103/1319-1683.68787] [Medline: 22022669]

2. Issenberg SB, McGaghie WC, Gordon DL, Symes S, Petrusa ER, Hart IR, et al. Effectiveness of a cardiology review course for internal medicine residents using simulation technology and deliberate practice. Teach Learn Med 2002;14(4):223-228. [doi: 10.1207/S15328015TLM1404_4] [Medline: 12395483]
3. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychological Review 1993;100(3):363-406. [doi: 10.1037/0033-295x.100.3.363]
4. Flanagan B, Nestel D, Joseph M. Making patient safety the focus: crisis resource management in the undergraduate curriculum. Med Educ 2004 Jan;38(1):56-66. [doi: 10.1111/j.1365-2923.2004.01701.x] [Medline: 14962027]
5. Barton J, Cybyk B, Drewry D, Frey T, Funk B, Hawthorne RC, et al. Use of a High-Fidelity UAS for Design, Testing, Training, and Mission Planning for Operation in Complex Environments. In: Wiley Encyclopedia of Operations Research and Management Science. Hoboken, NJ: John Wiley & Sons, Inc; Jan 14, 2011.
6. Okuda Y, Bryson EO, DeMaria S, Jacobson L, Quinones J, Shen B, et al. The utility of simulation in medical education: what is the evidence? Mt Sinai J Med 2009 Aug;76(4):330-343. [doi: 10.1002/msj.20127] [Medline: 19642147]
7. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. Acad Med 2003 Aug;78(8):783-788. [doi: 10.1097/00001888-200308000-00006] [Medline: 12915366]
8. Zendejas B, Wang AT, Brydges R, Hamstra SJ, Cook DA. Cost: the missing outcome in simulation-based medical education research: a systematic review. Surgery 2013 Feb;153(2):160-176. [doi: 10.1016/j.surg.2012.06.025] [Medline: 22884087]
9. Quick JA. Simulation Training in Trauma. Mo Med 2018;115(5):447-450 [FREE Full text] [Medline: 30385994]
10. Carron P, Trueb L, Yersin B. High-fidelity simulation in the nonmedical domain: practices and potential transferable competencies for the medical field. Adv Med Educ Pract 2011;2:149-155 [FREE Full text] [doi: 10.2147/AMEP.S19447] [Medline: 23745086]
11. Harrington CM, Kavanagh DO, Wright Ballester G, Wright Ballester A, Dicker P, Traynor O, et al. 360° Operative Videos: A Randomised Cross-Over Study Evaluating Attentiveness and Information Retention. J Surg Educ 2018;75(4):993-1000. [doi: 10.1016/j.jsurg.2017.10.010] [Medline: 29122571]
12. Pulijala Y, Ma M, Pears M, Peebles D, Ayoub A. Effectiveness of Immersive Virtual Reality in Surgical Training-A Randomized Control Trial. J Oral Maxillofac Surg 2018 May;76(5):1065-1072. [doi: 10.1016/j.joms.2017.10.002] [Medline: 29104028]
13. Buchman S, Miller C, Henderson D, Williams ER, Ray S. Interprofessional Students Learning to Save a Life through Cine-VR Simulation. EC Nursing and Healthcare 2020 Jul;22(11):2020.
14. Erickson M. Virtual reality system helps surgeons, reassures patients. Stanford Medicine. 2017 Jul 11. URL: https://med.stanford.edu/news/all-news/2017/07/virtual-reality-system-helps-surgeons-reassures-patients.html [accessed 2019-09-20]
15. Baker M. How VR is Revolutionizing the Way Future Doctors are Learning About Our Bodies. UCSF. 2017 Sep 18. URL: https://www.ucsf.edu/news/2017/09/408301/how-vr-revolutionizing-way-future-doctors-are-learning-about-our-bodies [accessed 2019-09-20]
16. Palter VN, Grantcharov TP. Individualized deliberate practice on a virtual reality simulator improves technical performance of surgical novices in the operating room: a randomized controlled trial. Ann Surg 2014 Mar;259(3):443-448. [doi: 10.1097/SLA.0000000000000254] [Medline: 24503910]
17. Seymour NE. VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. World J Surg 2008 Feb;32(2):182-188. [doi: 10.1007/s00268-007-9307-9] [Medline: 18060453]
18. Beverly EA, Love C, Love M, Williams E, Bowditch J. Virtual Reality Improves Healthcare Providers' and Administrators' Cultural Self-Efficacy and Diabetes Attitudes in Appalachian Ohio. JMIR Preprints Preprint posted online August 20, 2020. [FREE Full text] [doi: 10.2196/preprints.23708]
19. Buchman S, Henderson D. Interprofessional empathy and communication competency development in healthcare professions' curriculum through immersive virtual reality experiences. Journal of Interprofessional Education & Practice 2019 Jun;15:127-130. [doi: 10.1016/j.xjep.2019.03.010]
20. Mock C, Lormand J, Goosen J, Joshipura M, Peden M. Guidelines for essential trauma care. World Health Organization. 2004. URL: https://apps.who.int/iris/bitstream/handle/10665/42565/9241546409_eng.pdf?sequence=1 [accessed 2020-11-13]
21. No authors listed. Simulation in trauma education: beyond ATLS. Injury 2014 May;45(5):817-818. [doi: 10.1016/j.injury.2014.01.010] [Medline: 24502986]
22. LeBlanc VR. The effects of acute stress on performance: implications for health professions education. Acad Med 2009 Oct;84(10 Suppl):S25-S33. [doi: 10.1097/ACM.0b013e3181b37b8f] [Medline: 19907380]
23. Groombridge CJ, Kim Y, Maini A, Smit DV, Fitzgerald MC. Stress and decision-making in resuscitation: A systematic review. Resuscitation 2019 Nov;144:115-122 [FREE Full text] [doi: 10.1016/j.resuscitation.2019.09.023] [Medline: 31562904]
24. Ford K, Menchine M, Burner E, Arora S, Inaba K, Demetriades D, et al. Leadership and Teamwork in Trauma and Resuscitation. West J Emerg Med 2016 Sep;17(5):549-556 [FREE Full text] [doi: 10.5811/westjem.2016.7.29812] [Medline: 27625718]
25. GridLAB. 2020. URL: https://www.gridlab.org [accessed 2019-09-20]
26. Chandler P, Sweller J. Cognitive Load Theory and the Format of Instruction. Cognition and Instruction 1991 Dec;8(4):293-332. [doi: 10.1207/s1532690xci0804_2]

27. Kunkler K. The role of medical simulation: an overview. Int J Med Robot 2006 Sep;2(3):203-210. [doi: 10.1002/rcs.101] [Medline: 17520633]
28. Weller JM, Nestel D, Marshall SD, Brooks PM, Conn JJ. Simulation in clinical teaching and learning. Med J Aust 2012 May 21;196(9):594. [doi: 10.5694/mja10.11474] [Medline: 22621154]
29. Barfield W, Zeltzer D, Sheridan T, Slater M. Presence and performance within virtual environments. In: Barfield W, Furness TA, editors. Virtual environments and advanced interface design. New York, NY: Oxford University Press, Inc; Jun 1995:473-513.
30. Williams ER, Williams PS. How immersive 360-degree video of actual patient care can be used for VR medical professional education. 2017 Presented at: 1st Annual Virtual Reality and Healthcare Symposium; October 2017; Washington, DC.
31. Williams P, Chang K, Hutton H, Hansen J, Lerner A, Kroneberger L, et al. Preparing for Reality of the Clinic with Immersive Virtual Reality 360 Video/Audio. 2018 Presented at: APTA NEXT Conference; June 2018; Orlando, FL.
32. O'Sullivan B, Alam F, Matava C. Creating Low-Cost 360-Degree Virtual Reality Videos for Hospitals: A Technical Paper on the Dos and Don'ts. J Med Internet Res 2018 Jul 16;20(7):e239 [FREE Full text] [doi: 10.2196/jmir.9596] [Medline: 30012545]
33. Herault RC, Lincke A, Milrad M, Forsgärde E, Elmqvist C. Using 360-degrees interactive videos in patient trauma treatment education: design, development and evaluation aspects. Smart Learn. Environ 2018 Oct 19;5(1):1. [doi: 10.1186/s40561-018-0074-x]

Abbreviations

ATLS: Advanced Trauma Life Support
cineVR: 360-degree video and audio
CLT: Cognitive Load Theory
ED: emergency department
EMS: emergency medical service
GRID: Game Research and Immersive Design
HIPAA: Health Insurance Portability and Accountability Act
IRB: institutional review board
VR: virtual reality

Edited by G Eysenbach; submitted 10.07.20; peer-reviewed by H Yaremych, S Tierney; comments to author 28.07.20; revised version received 30.10.20; accepted 03.11.20; published 16.12.20.

Please cite as:
Patel D, Hawkins J, Chehab LZ, Martin-Tuite P, Feler J, Tan A, Alpers BS, Pink S, Wang J, Freise J, Kim P, Peabody C, Bowditch J, Williams ER, Sammann A
Developing Virtual Reality Trauma Training Experiences Using 360-Degree Video: Tutorial
J Med Internet Res 2020;22(12):e22420
URL: http://www.jmir.org/2020/12/e22420/
doi:10.2196/22420
PMID:33325836

Original Paper

Mobile Texting and Lay Health Supporters to Improve Schizophrenia Care in a Resource-Poor Community in Rural China (LEAN Trial): Randomized Controlled Trial Extended Implementation

Yiyuan Cai[1,2*], MM; Wenjie Gong[3*], PhD; Hua He[4], PhD; James P Hughes[5], PhD; Jane Simoni[6], PhD; Shuiyuan Xiao[3], PhD; Stephen Gloyd[7], MD, MPH; Meijuan Lin[1], MM; Xinlei Deng[1], BA; Zichao Liang[1], BA; Wenjun He[1], MM; Bofeng Dai[1], MM; Jing Liao[1], PhD; Yuantao Hao[1], PhD; Dong (Roman) Xu[8*], PhD

[1]School of Public Health, Sun Yat-sen University, Guangzhou, China
[2]School of Public Health, Guizhou Medical University, Guiyang, China
[3]Xiangya School of Public Health, Central South University, Changsha, China
[4]School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, United States
[5]Department of Biostatistics, University of Washington, Seattle, WA, United States
[6]Department of Psychology, University of Washington, Seattle, WA, United States
[7]Department of Global Health, University of Washington, Seattle, WA, United States
[8]Acacia Lab for Health Systems Strengthening and Department of Health Management, School of Health Management, Southern Medical University, Guangzhou, China
[*]these authors contributed equally

Corresponding Author:
Yuantao Hao, PhD
School of Public Health
Sun Yat-sen University
No 74 Zhongshan Road 2, Yuexiu District
Guangzhou
China
Phone: 86 020 87335523
Email: haoyt@mail.sysu.edu.cn

Abstract

Background: Schizophrenia is a severe and disabling condition that presents a dire health equity challenge. Our initial 6-month trial (previously reported) using mobile texting and lay health supporters, called LEAN, significantly improved medication adherence from 0.48 to 0.61 (adjusted mean 0.11, 95% CI 0.03 to 0.20, *P*=.007) for adults with schizophrenia living in a resource-poor village in rural China.

Objective: We explored the effectiveness of our texting program in improving participants' medication adherence, functioning, and symptoms in an extended implementation of the intervention after its initial phase.

Methods: In an approximated stepped-wedge wait-list design randomized controlled trial, 277 community-dwelling villagers with schizophrenia were assigned 1:1 in phase 1 into intervention and wait-list control groups. The intervention group received (1) lay health supporters (medication or care supervisors), (2) e-platform (mobile-texting reminders and education message) access, (3) a token gift for positive behavioral changes, and (4) integration with the existing government community-mental health program (the 686 Program) while the wait-listed control group initially only received the 686 Program. Subsequently (in the extended period), both groups received the LEAN intervention plus the 686 Program. The primary outcome was antipsychotic medication adherence (percentage of dosages taken over the past month assessed by unannounced home-based pill counts). The secondary outcomes were symptoms measured during visits to 686 Program psychiatrists using the Clinical Global Impression scale for schizophrenia and functioning measured by trained student assessors using the World Health Organization Disability Assessment Schedule 2.0. Other outcomes included data routinely collected in the 686 Program system (refill records, rehospitalization due to schizophrenia, death for any reason, suicide, wandering, and violent behaviors). We used intention-to-treat analysis and missing data were imputed. A generalized estimating equation model was used to assess program effects on antipsychotics medication adherence, symptoms, and functioning.

Results: Antipsychotics medication adherence improved from 0.48 in the control period to 0.58 in the extended intervention period (adjusted mean difference 0.11, 95% CI 0.04 to 0.19; *P*=.004). We also noted an improvement in symptoms (adjusted mean difference –0.26, 95% CI –0.50 to –0.02; *P*=.04; Cohen *d* effect size 0.20) and a reduction in rehospitalization (0.37, 95% CI 0.18 to 0.76; *P*=.007; number-needed-to-treat 8.05, 95% CI 4.61 to 21.41). There was no improvement in functioning (adjusted mean difference 0.02, 95% CI –0.01 to 0.06; *P*=.18; Cohen *d* effect size 0.04).

Conclusions: In an extended implementation, our intervention featuring mobile texting messages and lay health workers in a resource-poor community setting was more effective than the 686 Program alone in improving medication adherence, improving symptoms, and reducing rehospitalization.

Trial Registration: Chinese Clinical Trial Registry; ChiCTR-ICR-15006053 https://tinyurl.com/y5hk8vng

(J Med Internet Res 2020;22(12):e22631) doi:10.2196/22631

KEYWORDS

medication adherence; mobile texting; lay health worker; resource-poor community; primary health care; quality of care; mHealth; schizophrenia

Introduction

Affecting 0.4% of the population [1], schizophrenia is a common and disabling condition that presents a dire health equity challenge. Schizophrenia can often be effectively controlled with life-long antipsychotic medications [2]. No access or poor adherence to treatment has serious consequences for the patient, family, and society, leading to a higher risk of worsening symptoms, repeated and prolonged hospitalizations, suicide, aggressive conduct, poor quality of life, and reduced functioning [3]. Nearly half of people with schizophrenia in low-and-middle-income countries take less than 70% of prescribed doses [4]. To address these challenges, the World Health Organization Mental Health Gap Action Program recommended a community-based approach whereby community health workers play an active role in schizophrenia assessment, treatment, rehabilitation, and follow-up [5]. Along the same lines, China rolled out its National Continuing Management and Intervention Program for Psychoses, known as the 686 Program, in 2005 [6,7]. The program covered 5,810,000 people with psychoses across China in 2017 [8] and has become part of China's "integrated public mental health service [8,9]." However, even though the program provides free medication to individuals with low-incomes, less than 40% of program enrollees routinely adhered to their antipsychotic dosage [7].

Mobile texting or SMS has been successfully used to improve schizophrenia—particularly as a way to improve medication adherence—however, evidence of effectiveness was based on short-term interventions (mostly less than 6 months) or short-term follow-ups and in high-income settings [10-14]. It is important to explore the long-term effect of mobile health interventions because, while longer implementations are able to better shape new behavior and allow sufficient time for effects on functioning and symptoms to occur [15], participants eventually lose interest in program participation. For real-world practicality and sustainability, it is thus critical to examine the effects of the extended implementation of mobile texting interventions.

Initially, we conducted a wait-list randomized controlled study that tested the effectiveness of LEAN, a community-based solution that consists of mobile-texting for medication reminders, relapse monitoring, and education and also a lay health supporter selected from the patient families to help with patient care [16,17]. In this study, we hypothesized that the extended implementation of LEAN (1) would maintain the effects on medication adherence shown in the initial phase, and (2) would begin to show effects on symptoms and functioning that were initially absent.

Methods

Protocol

Details of the study design (including sample size calculation and participant randomization), methods, and analysis plan have previously been published as a study protocol [16].

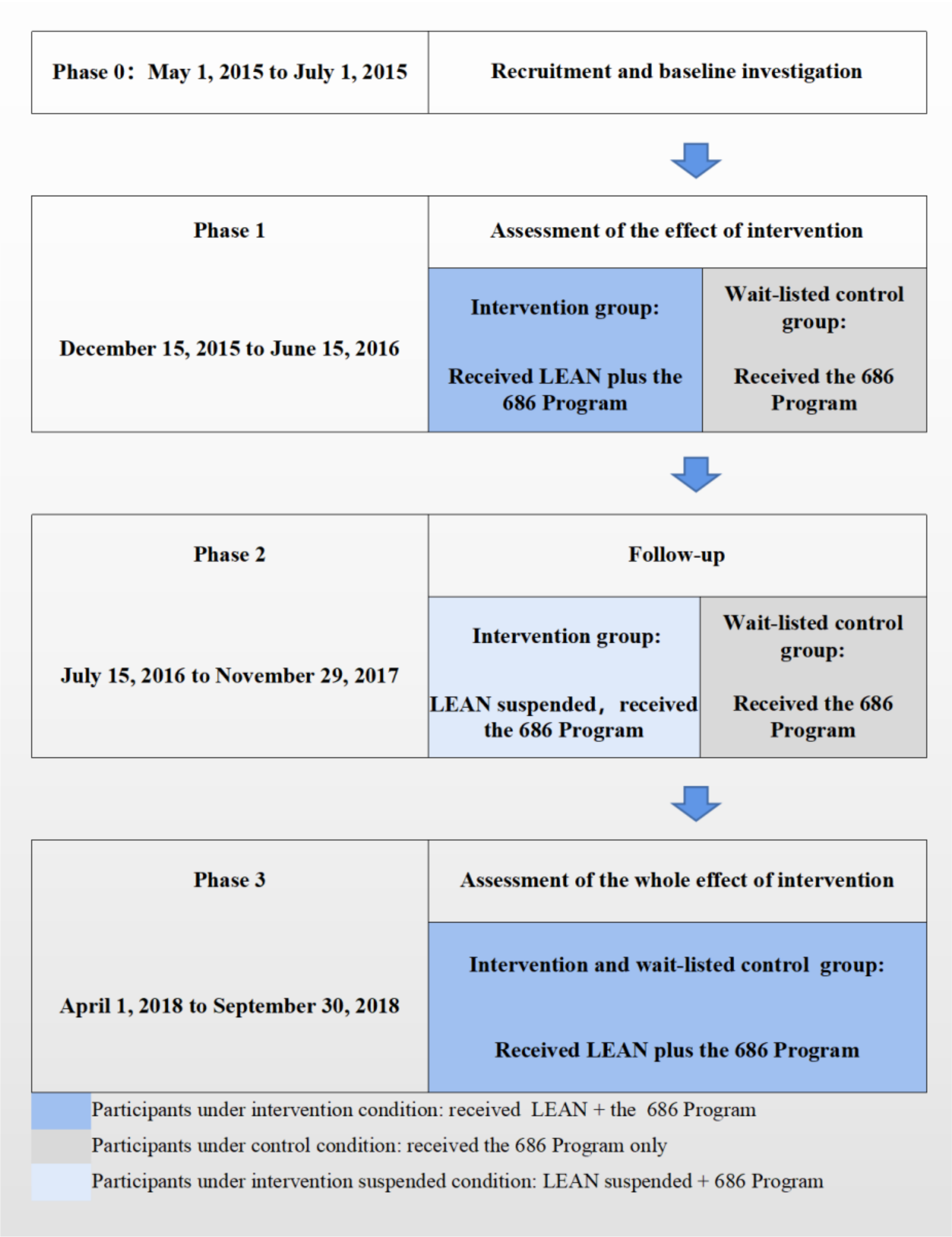
Data

Data will be freely available upon request after the publication of this manuscript.

Trial Design

Because of the initial study design [16], in which we employed a wait-listed control group that would receive the intervention once it was proven to be effective, it allowed a 2-period intervention trial similar to a stepped-wedge design trial [18]. The project was executed in 3 phases (Figure 1): phase 1 was from December 15, 2015 to June 15, 2016 during which only the intervention group received LEAN for 6 months; phase 2 (blank phase) consisted of a data analysis period from July 15, 2016 to November 29, 2017 during which neither group received LEAN; and phase 3 was the formal extended trial from April 1, 2018 to September 30, 2018 during which both of the original groups received LEAN for 6 months. The results of phase 1 have been published [17].

Figure 1. Trial design.



Setting and Participants

The trial was conducted in the communities of 9 rural townships of Liuyang municipality (population 356,900) in the Hunan province of China. The patient participants of the program were required to (1) be community-dwelling, (2) be enrollees of the 686 Program, (3) have a primary diagnosis of schizophrenia according to the International Statistical Classification of Diseases, Tenth Revision [19] (diagnosis reconfirmed by the 686 Program psychiatrists while they were enrolled), (4) be on oral psychotropic medication, and (5) be residents of 1 of the 9 rural townships. Patients were excluded if they (1) were hospitalized due to schizophrenia at the time of recruitment (our approach was community-based), or (2) had missed 3 immediate consecutive past drug refills (in this case, they had de facto dropped out of the 686 Program), or (3) were physically incapable of using voice and text messaging (hearing or vision impairment prevented the use of our intervention) [16]. The trial participants were selected by simple random sampling from the 686 Program registry that included almost all known residents of Liuyang diagnosed with schizophrenia.

Procedures

The development process of our intervention LEAN was described in detail in our earlier paper with phase 1 results [17]. In summary, we used the Health Belief Theory [20-22] and empirical evidence to guide the development of the prototype LEAN intervention, which was refined and finalized through trial and error in the pilot. LEAN includes 4 elements: (1) lay health supporters (often a designated family member to monitor patient medication, side effects, and relapses, and urgent care), (2) e-platform (a texting system for medication reminders, health education, and relapse monitoring), (3) token gifts to encourage behavioral improvement, and (4) integration of the texting with the existing health system to enable collaborative care. Following the same procedure [16], 4 masters' students in public health produced educational and reminder text messages, which were reviewed by a senior psychiatrist. Frequency and timing of the messages were determined after multiple rounds of consultation with the patients and their families: medication reminders at 7 PM daily, educational messages at 9 AM every other day (these had been sent every day in phase 1), and relapse-monitoring messages monthly. In phase 3, all patients and their lay health supports received the message if they had a cellphone; however, in phase 1, only patients and their lay health supports in the intervention group received the message.

In phase 1, trainers had used one-on-one *demonstration-imitation* to teach patients and lay health supporters in the intervention arm how to use a cellphone to read and reply to the texted messages [23]. Lay health supporters were also trained how to read and reply to relapse-monitoring messages in a group training session in each program township. Lay health supporters were trained to remind the patient to take their medication and report signs of relapse. For the extended intervention, during phase 2, program staff performed the same one-on-one training again during their visits to the patients' homes. Before the extended intervention was formally implemented, we conducted a 1-week pilot by sending messages to guide participants to practice reading and replying.

Both the initial intervention and control groups received the 686 Program. In Liuyang, 2 internist-transformed psychiatrists along with several staff members traveled with medication supplies to each township's health center at a fixed date every 2 months, providing patients with a brief consultation and medication adjustment and refills. The township's mental health administrators (generally public health professionals) informed the patients and their family by phone to meet the traveling psychiatrists and worked with the village doctors (paramedics with rudimentary medical training) to provide yearly physical exams, assessment of risk level, ≥4 home visits throughout the year, health education, and urgent care.

Outcomes

We continued to track the outcomes specified and detailed in our published protocol [16]. We collected the primary and secondary outcomes in 2 home visits conducted in phase 2 from November 24 to 29, 2017, and from December 24 to 29, 2017, respectively. We again conducted home-based visits from July 19 to 24 and August 19 to 24 in 2018 in phase 3. The primary outcome was antipsychotic medication adherence using the proportion of doses taken in the past month (*number of the first count – number of the second count + number of additional pills obtained – number of pills discarded*) / (*number of pills prescribed*) assessed by unannounced home-based pill-counts. If 2 or more kinds of antipsychotic medications were taken, we used the combined total number of tablets of the medications required by the prescription as the denominator and the actual number of consumed pills as numerator to calculate adherence. Two counts (30 days apart) were necessary to calculate the pills taken over 1 month: the first at the beginning and the second at the end of the last month of each phase. The numbers of pills prescribed were extracted from the 686 Program system. Assessors followed a standard protocol to ask family members and patients about the number of pills additionally purchased and deliverable discarded. The count was considered unannounced since the participants consented to the count but were unaware of the specific timing of the count. Besides, medication refill adherence was assessed (*number of refills required / number of refills conducted over past 6 months*), and other self-reported adherence measures were assessed with the Brief Adherence Rating Scale (BARS) [24] and the Drug Attitude Inventory-10 (DAI-10) [25,26] during each home-based interview. Secondary outcomes were patient functioning and patient symptoms assessed with the World Health Organization Disability Assessment Schedule 2.0 (WHO DAS) [27,28] and the Clinical Global Impression in Schizophrenia measure (CGI) [29], respectively.

Trained public health, nursing, and health management students performed the home-based visits. If the patients could communicate with the assessors, the results of BARS, DAI-10, and WHO DAS were reported by the patients, otherwise, the BARS and WHO DAS were reported by their lay health supports. As part of the process evaluation, the students also assessed user experience at the home visits with questionnaires. When 686 Program psychiatrists conducted routine visits to each town at the end of each phase, they performed symptom evaluations using CGI.

We also collected baseline information on medication side effects, substance use, and family supervision when taking medicine since they are strong predictors of adherence that have been empirically suggested by other studies [30]. Medication side effects assessed with the self-reported Glasgow Antipsychotic Side-effect Scale [31]; smoking and alcohol use were also self-reported.

We also extracted routinely collected outcomes from the 686 Program system including rehospitalization due to schizophrenia, death for any reason, suicide, wandering, and violence for the second intervention period, which were verified during the home-base interview. These outcomes are recorded by mental health administrators when they routinely interviewed participants at home or by phone every 3 months. We also captured the frequency that text recipients responded through the texting system log and obtained program-cost information for various program operation channels.

Statistical Methods

Analyses for phase 3 approximated those for a stepped-wedge randomized controlled design. In a stepped-wedge design, all groups received randomized orders to enter the intervention in sequence. Thus, we no longer used the term intervention or control groups but intervention and control periods instead [32]. The stepped-wedge design maintained the nature of randomization and had the advantage of not only using 2 groups for cross-sectional analysis but also before-and-after comparison of the same group for longitudinal analysis.

We implemented a generalized estimating equation model [33-36] (gee, version 4.8; R, version 3.5.3) in all analyses and used *link identity* for continuous outcomes and logit for binary outcomes and assumed an exchangeable correlation structure. The main results of phase 3 were compared in 2 periods (ie, the intervention and control periods). Statistical significance was set at $P<.05$.

The analysis of primary outcome (adherence) adjusted for the empirically suggested and prespecified baseline predictors, which were the same as those in phase 1 analysis included baseline adherence, the overall severity of illness, negative symptoms, functioning, substance use, medication side effects, and family supervision. Analyses of the secondary outcomes (WHO DAS and CGI severity of illness and CGI degree of change) were adjusted by their baselines. We used the same analysis models and the covariate adjustment for 2 prespecified subgroups—a group demonstrating baseline nonadherence (missing any of the previous 6 refills was considered nonadherence) and a group with low baseline functioning (cut-off ≤0.22). In extended analyses, we also explored whether the phase 3 intervention led to additional improvement among people who received the intervention in phase 1.

We performed several sensitivity analyses. We compared the results of the program effects on adherence, functioning, and symptoms with a raw analysis, the unadjusted analysis with data imputation for the missing data, and adjusted analysis with covariates and data imputation for the missing data. Meanwhile, for phase 3 analysis, since no one received the intervention during phase 2, we performed a sensitivity analysis by including phase 2 data, coding the intervention group for this phase as 1 and the wait-listed control group as 0. We analyzed antipsychotic medication adherence as a continuous variable, and we also explored the sensitivity of dichotomizing adherence cut-off points of 0.70, 0.80, and 0.90.

We used intention-to-treat analysis [37] for all participants for primary and secondary outcomes. We also conducted a per-protocol analysis. The fully conditional specification multiple imputation method was used to impute missing data [38]. To enable cross-study comparisons, we calculated the program effect size using Cohen *d* [39].

Ethics and Dissemination

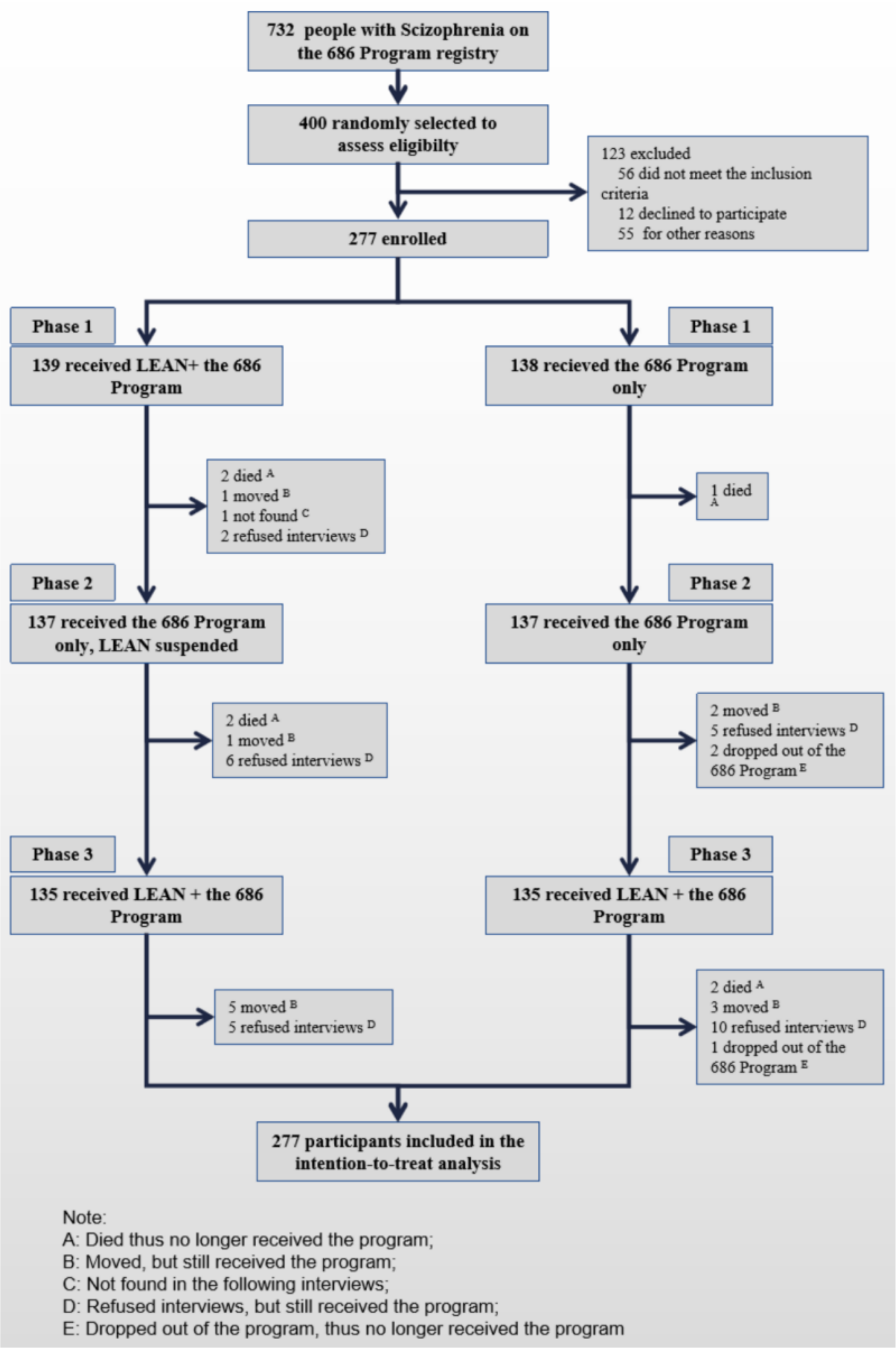
The study obtained institution review board approval from the University of Washington (49464 G) and Central South University (CTXY-150002-6). All patient participants and their lay health supporters provided written informed consent.

Results

Participants

In phase 1, there were 277 patient participants (intervention, n=139; wait-listed control group, n=138) (Figure 2). Due to the closed-cohort design, participants assessed in different periods remained the same. The baseline characteristics were well balanced between the 2 groups [17]. There were 244 participates successfully followed in phase 3. In phase 1, each patient participant in the intervention group had a designated lay health supporter, and in phase 3, the lay health supporters in the original intervention group continued to take the same responsibility. Each of the participants in the original control group also received a lay health supporter. The supporters were mainly family members (129/138), and the rest of the other supporters were community volunteers or village doctors.

Figure 2. Participant flowchart.



Note:
A: Died thus no longer received the program;
B: Moved, but still received the program;
C: Not found in the following interviews;
D: Refused interviews, but still received the program;
E: Dropped out of the program, thus no longer received the program

Retention and Outcomes

At the end of phase 3, we captured the information of 271/277 (97.8%) participants for medication refill adherence, 164/277 (59.2%) for pill-count adherence, 231/277 (83.4%) for functioning (WHO DAS), and 247/277 (89.17%) for symptoms. Analyses of patterns of missing data and the results of multiple imputation are presented in Multimedia Appendix 1.

Adherence

In phase 3, adherence increased from 0.48 in the control period (ie, control arm in phase 1) to 0.58 in the intervention periods (ie, intervention arm in phase 1 and both arms in phase 3). Model-based adjusted analysis showed the increase in antipsychotic adherence remained statistically significant (adjusted mean difference 0.11, 95% CI 0.04 to 0.19; *P*=.004; Cohen *d* effect size 0.28; Table 1). Subgroup analyses suggested that the participants with better adherence and poor functioning at baseline presented better adherence at the end of phase 3 (Figure 3). Distribution of adherence between the intervention and the control period are shown in Multimedia Appendix 2, and the reasons for zero adherence are explored in Multimedia Appendix 3.

Table 1. Results.

Measures	Intervention periods, mean (SD)	Control periods, mean (SD)	Mean difference (95% CI)	*P* value
Adherence outcomes				
Pill-count assessment[a]	0.58 (0.38)	0.48 (0.35)	0.11 (0.04, 0.19)	.004
Other				
Medication refill assessment	0.77 (0.35)	0.76 (0.34)	–0.01 (–0.06, 0.04)	.70
DAI-10[b]	0.67 (0.23)	0.67 (0.22)	0.01 (–0.04, 0.06)	.73
BARS[c]	0.71 (0.20)	0.68 (0.23)	0.04 (–0.01, 0.08)	.13
Secondary outcomes				
WHO[d] DAS[a,e]	0.15 (0.19)	0.15 (0.19)	0.02 (–0.01, 0.06)	.19
CGI[f] severity of illness[a,g]	2.51 (1.23)	2.76 (1.24)	–0.26 (–0.50, 0.02)	.04
Negative	2.41 (1.17)	2.98 (1.43)	–0.47 (–0.76, 0.19)	.001
Positive	2.42 (1.19)	2.67 (1.55)	–0.47 (–0.76, 0.17)	.002
Depression	1.9 (1.01)	2.11 (1.26)	–0.21 (–0.44, 0.01)	.07
Cognition	2.44 (1.19)	2.85 (1.44)	–0.50 (–0.79, 0.21)	<.001
CGI degree of change[a,h]	3.12 (1.16)	3.02 (1.08)	0.10 (0.00, 0.25)	.05
686 Program				
Rehospitalization due to schizophrenia	43 (11.1)	25 (19.5)	0.58 (0.37, 0.91)	.02
Death[i]	0 (1.0)	3 (2.2)	N/A	N/A
Suicide[i]	0 (0.0)	1 (0.0)	N/A	N/A
Wandering	7 (1.7)	2 (1.5)	0.98 (0.23, 4.13)	.98
Hurting people or smashing objects	2 (0.4)	6 (4.5)	0.11 (0.02, 0.54)	.006
Getting into trouble	0 (0.0)	0 (0.0)	0.11 (0.04, 0.19)	N/A
Self-harm	0 (0.0)	0 (0.0)	–0.01 (–0.06, 0.04)	N/A

[a] Adjusted for baseline covariates with imputation for missing data.

[b] DAI: Drug Attitude Inventory. The score was rescaled as 0 to 1.

[c] BARS: Brief Adherence Rating Scale.

[d] WHO: World Health Organization.

[e] DAS: Disability Assessment Schedule (range 0 to 1).

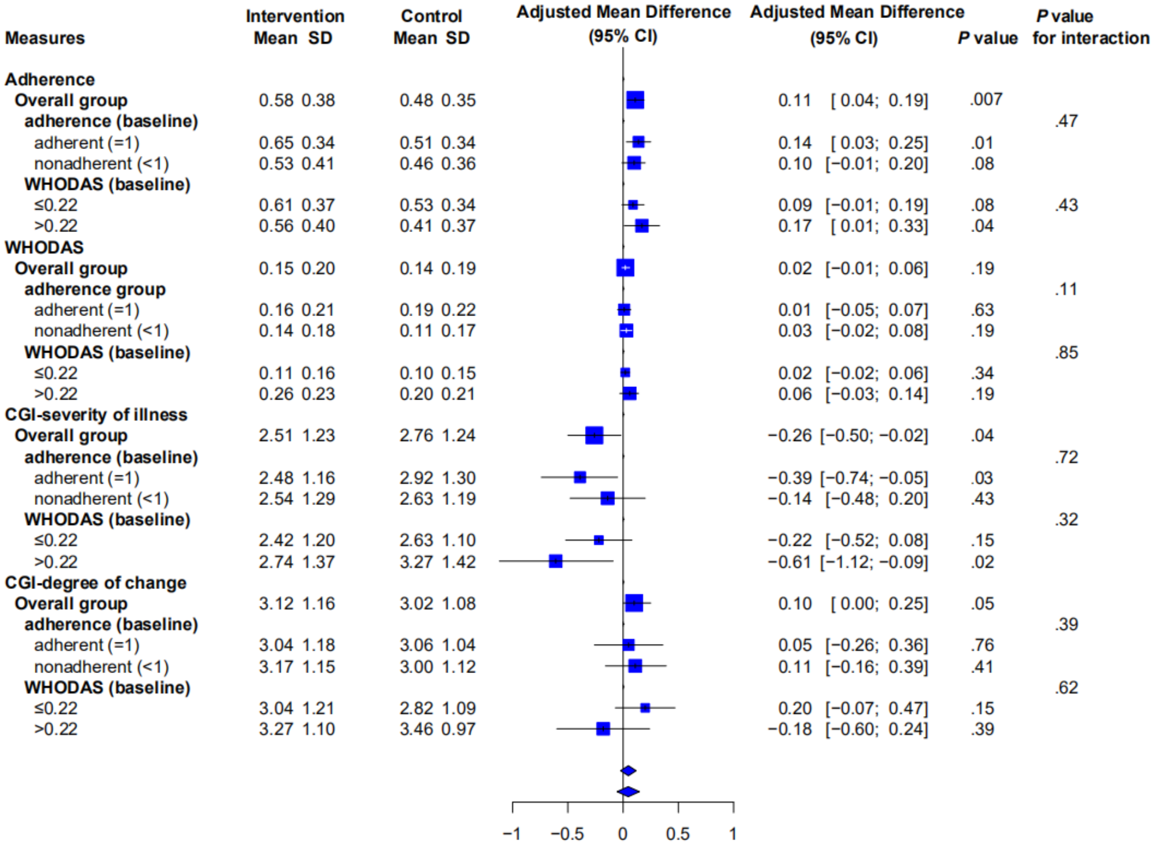
[f] CGI: Clinical Global Impression.

[g] Higher scores indicate worse symptoms (range 1 to 7).

[h] Higher scores indicate less change (range 1 to 7).

[i] The numbers are for phase 3 only.

Figure 3. Subgroup analysis. CGI: Clinical Global Impression; WHODAS: World Health Organization Disability Assessment Schedule.



Measures	Intervention Mean	Intervention SD	Control Mean	Control SD	Adjusted Mean Difference (95% CI)	*P* value	*P* value for interaction
Adherence							
Overall group	0.58	0.38	0.48	0.35	0.11 [0.04; 0.19]	.007	
adherence (baseline)							.47
adherent (=1)	0.65	0.34	0.51	0.34	0.14 [0.03; 0.25]	.01	
nonadherent (<1)	0.53	0.41	0.46	0.36	0.10 [−0.01; 0.20]	.08	
WHODAS (baseline)							
≤0.22	0.61	0.37	0.53	0.34	0.09 [−0.01; 0.19]	.08	.43
>0.22	0.56	0.40	0.41	0.37	0.17 [0.01; 0.33]	.04	
WHODAS							
Overall group	0.15	0.20	0.14	0.19	0.02 [−0.01; 0.06]	.19	
adherence group							.11
adherent (=1)	0.16	0.21	0.19	0.22	0.01 [−0.05; 0.07]	.63	
nonadherent (<1)	0.14	0.18	0.11	0.17	0.03 [−0.02; 0.08]	.19	
WHODAS (baseline)							.85
≤0.22	0.11	0.16	0.10	0.15	0.02 [−0.02; 0.06]	.34	
>0.22	0.26	0.23	0.20	0.21	0.06 [−0.03; 0.14]	.19	
CGI-severity of illness							
Overall group	2.51	1.23	2.76	1.24	−0.26 [−0.50; −0.02]	.04	
adherence (baseline)							.72
adherent (=1)	2.48	1.16	2.92	1.30	−0.39 [−0.74; −0.05]	.03	
nonadherent (<1)	2.54	1.29	2.63	1.19	−0.14 [−0.48; 0.20]	.43	
WHODAS (baseline)							.32
≤0.22	2.42	1.20	2.63	1.10	−0.22 [−0.52; 0.08]	.15	
>0.22	2.74	1.37	3.27	1.42	−0.61 [−1.12; −0.09]	.02	
CGI-degree of change							
Overall group	3.12	1.16	3.02	1.08	0.10 [0.00; 0.25]	.05	
adherence (baseline)							.39
adherent (=1)	3.04	1.18	3.06	1.04	0.05 [−0.26; 0.36]	.76	
nonadherent (<1)	3.17	1.15	3.00	1.12	0.11 [−0.16; 0.39]	.41	
WHODAS (baseline)							.62
≤0.22	3.04	1.21	2.82	1.09	0.20 [−0.07; 0.47]	.15	
>0.22	3.27	1.10	3.46	0.97	−0.18 [−0.60; 0.24]	.39	

Functioning

There were 31.8% (75/236) WHO DAS questionnaires completed by lay health supporters on behalf of patient participants. Interviewees were the patient’s parents (31/75), spouses (28/75), offspring (8/75), siblings (6/75), and others (2/75) (Multimedia Appendix 4). At the end of phase 3, we observed no improvement in functioning (Table 1). For the raw analysis, the WHO DAS score changed from 0.14 under the control period to 0.15 under the intervention periods; our model-based analysis showed no difference (adjusted mean difference 0.02, 95% CI –0.01 to 0.06; *P*=.18; Cohen *d* effect size 0.04; Table 1). Neither did we note any improvement in the severity of functioning for the prespecified subgroups (Figure 3).

Symptoms

The severity of illness decreased (CGI score 2.76 in the control period to 2.51 in the intervention periods (model-based adjusted mean difference –0.26, 95% CI –0.50 to –0.02; *P*=.04; Cohen *d* effect size 0.20). All domains of the CGI (negative: *P*=.001, positive: *P*=.002, cognition: *P*<.001) except depression (*P*=.07) showed significant improvements. Subgroup analyses suggested that the baseline adherence group and poor functioning group presented symptom improvement at the end of phase 3 as well (Figure 3).

Other Outcomes

The incidence of rehospitalization due to schizophrenia was 19.5% (25/128) in the control period versus 11.1% (43/387) in the intervention periods (relative risk 0.58, 95% CI 0.37 to 0.91, *P*=.02, number-needed-to-treat 12.25 95%CI 6.36 to 166.49; Table 1).

There was a significant difference in violence (hurting people or smashing objects) (*P*=.006). However, there were no significant differences in wandering (*P*=.98) and in other outcomes captured in the 686 Program. Eight patients died during the entire follow-up including 2 patients who committed suicide (all-cause mortality 28.88/1000 person-year).

Extended Analyses and Sensitivity Analyses

Through our extended analyses, we found that the phase 3 intervention did not result in any additional improvement in adherence (mean difference –0.034, 95% CI –0.141 to 0.073; *P*=.531) and functioning (mean difference –0.84, 95% CI –0.626 to 0.458; *P*=.084) among participants who received the intervention in phase 1. However, their symptoms were further improved in phase 3 (mean difference –0.407, 95% CI –0.713 to –0.101; *P*=.009).

Sensitivity analyses (see Multimedia Appendix 5) showed that the results were not sensitive to the different analytical methods used. The effects of the alternative analytical approaches were almost identical to that of our primary analysis, probably mainly due to the possible reasons that (1) missing data were possibly missing at random, and (2) few participant-lay health supporter pairs did not receive the intervention (14/270, 5.2%), and the exclusion of these samples had little effect on the results of the analysis.

Process Indicators

Process indicators including the process of texting, user experiences, and cost information are presented in Multimedia Appendix 6. We also present example text messages in Multimedia Appendix 6.

Discussion

Principal Results

In this stepped wedge–like randomized controlled trial, we found that a long-term intervention of texting for individuals with schizophrenia and their lay health supporters improved pill count–based adherence to antipsychotic medication from 0.48 during the control period to 0.58 during the intervention period, reduced CGI severity of illness score from 2.76 to 2.51, and reduced rehospitalization from 19.7% to 11.8%.

As of December 1, 2019, we identified 6 randomized controlled trials of mobile texting to improve care for people with schizophrenia [10-14,40]. The 4 trials that had less than 6 months of intervention all presented positive effects of the intervention [10-14,40]. However, two 12-month interventions showed no difference between the intervention and control groups [13,40]. All of these previous trials were conducted in high-income countries. In the LEAN trial, mobile texting improved adherence with a Cohen *d* effect size of 0.28, smaller than that of Montes et al’s study [12], which used patient self-reported adherence measured with the Morisky Green Adherence Questionnaire, but larger than that of Beebe et al’s study [14], which also used in-home pill-count adherence (Multimedia Appendix 7). We suspect our texting of both the patients and their lay health supporters (mostly family members) was a major contributor to LEAN’s relative superiority in the long-term implementation. Our texted medication reminders, health education, and relapse monitoring may have provided tools and cues for the family to take more active and timely action in supervising medications. Despite LEAN’s effect, we should also note that there seemed to be a decline in improving adherence in the long-term intervention (0.48 in control to 0.61 in phase 1 [17] versus 0.48 to 0.58 in phase 3). Our analyses suggested two possible explanations. First, because these individuals with schizophrenia lack awareness of the severity of the disease, more patients stopped taking their medication in phase 3. This may be caused by the intervention not helping these participants to develop a belief and hence sustainable behavior to keep taking medicine. Second, we observed a sign of participants’ fatigue toward the program, as both participant's satisfaction toward LEAN and the family response rates decreased from phase 1 to phase 3, (Multimedia Appendix 6, Table S2). Välimäki et al [40] also reported fatigue in their study. Therefore, to sustain mobile health interventions for people with schizophrenia, timely adjustments to the program may be required to counter the possible decline in participants’ satisfaction.

We expected that increased antipsychotic medication adherence would translate into improved symptoms and functioning in the long-term. However, a possible ceiling effect may have prevented further improvements in functioning. Likely due to the 686 Program that both the intervention and control groups received, LEAN participants at baseline had much better functioning (WHO DAS score 0.18) than other groups of people with schizophrenia of similar ethnicity and culture have shown (WHO DAS score ranging from 0.29 [41] to 0.64 [42]). However, despite the overall mild symptoms of LEAN participants, an improvement in symptoms (reduction of CGI severity of illness score from 2.76 to 2.51) was still observed at the end of phase 3 but not in phase 1. This may suggest that over an extended time the enhancement in medication adherence may gradually result in symptom reduction. Furthermore, LEAN’s effect on symptoms was in concordance with the observed large reduction in rehospitalization due to schizophrenia.

In extended analyses, we found that the phase 3 intervention led to additional improvement in symptoms but not in adherence and functioning among people with schizophrenia who received the intervention in phase 1. We suspect that further improvements in adherence and functioning were difficult due to a ceiling effect. All participants had received the 686 Program, and thus a basic level of function had been maintained. However, changes in symptoms may be more sensitive to changes in adherence over a long time.

Limitations

We should note several limitations of the LEAN trial. First, we did not take a systematic approach in assessing the reasons for the participants discontinuing and continuing antipsychotic medications. Future trials may consider using scales such as the Antipsychotic Discontinuation Questionnaire [43-45]. Second, we obtained the incidence of relapse (defined as the marked occurrence of symptoms) by in-person interviewing of the family carers in phase 1 and by self-reported relapse from people with schizophrenia or their lay health supporters through their response to our relapse monitoring messages in phase 3. We felt more objective and defined measures for relapse should be used for future studies. Third, the blank phase (phase 2) when neither the intervention nor the control groups received LEAN created an aberration to the standard stepped-wedge design, although we included this phase as a covariate in the secondary model. Fourth, in phase 3, 34.1% (92/270) patient participants and 17.8% (48/270) lay health supporters were not equipped with a phone; however, there were a total of 5.2% (14/270) pairs (of patient participants and lay health supporters) not equipped with a phone, either not having a phone or not having replaced a damaged, malfunctioning, or lost phone in phase 1. Fifth, our intervention was an integrated package that combined 4 program elements, of which 2 critical components were the use of lay health supporters and text messages. We are not able to determine the individual contribution of either component to the program’s effect. In our study, we did not systematically collect information about the depth of the involvement of the lay health supporters. We were unable to quantitatively isolate the effect of the role played by the lay health supporters in the overall intervention package. However, we did track the satisfaction level of the supporters and found out that there was a decreasing trend in overall satisfaction near the end of our intervention. It showed that 80.0% (100/125) lay health supporters surveyed were satisfied with the intervention, and 78.4% (98/125) expressed their willingness to remain in phase

3 (see Multimedia Appendix 6, Table S2). Texting system records showed that 47.3% (105/222) of family members responded to the texted reminders. Future studies may consider multiphase optimization strategy and sequential multiple assignment randomized trial designs to fully address those issues [46,47]. Besides, LEAN used daily text reminders but the optimal frequency was not tested. The use of multiphase optimization and sequential multiple assignment randomized trial designs can also address this issue well. Finally, although we used a rigorous method of adherence assessment through unannounced home-based pill-counts, we could not completely address the issue of participants deliberately discarding pills. However, this should not affect our program impact evaluation as the behavior may occur in both the control and intervention periods. Also, we believe our method of assessing adherence is best suited for this study—other methods, such as an electronic cap, may remind people to take medications and thus interfere in the unbiased assessment of adherence [48].

Due to the complexity of analysis and the different hypotheses tested, we decided to cover the details of the results of phase 2 in a separate manuscript that will primarily explore the maintenance of the effects after the withdrawal of the intervention.

Comparison With Prior Work

Universal health coverage is not attainable without the coverage for mental health. China's 686 Program provides valuable experiences for other low-resource settings to develop community-based care at scale. LEAN further enabled and motivated family carers and patients with an easy-to-implement and low-cost texting system. The use of mobile texting for mental health is perhaps no longer considered novel. However, few trials have attempted to test its effectiveness in long-term implementation. LEAN spanned over 2 years with 1 year of dedicated texting and demonstrated overall effectiveness in improving medication adherence, symptoms, and rehospitalization. Although LEAN was built upon China's 686 Program, many elements of LEAN should be useful for other low-resource settings with or without an existing community-based program. Implications of LEAN results to other low-resources settings should note the following. First, the overall adherence to antipsychotic medication at the endpoint remained low at 0.57 (SD 0.40) despite free medication from the 686 Program and the LEAN efforts. Long-acting injectable antipsychotic medications, in treating schizophrenia, were cost-effective [49] and reduced relapse [50] but have not been widely accepted by clinicians and families with schizophrenia in China [51]. Second, formal health workers in China and other low-resource settings are often overextended, so task-shifting is important such as engaging the self-motivated family members in monitoring medication, side effects, and relapse. LEAN provides a relatively low-cost approach to assist the family carers in their roles. In our study, the majority of lay health supporters were family members who lived with the patient participants (258/277); they reminded them to take their medication and monitored signs of side-effects and relapses. Third, SMS played a very important role in connecting the patient participants, their families, medical staff, and the health care system. It created a way of communication and timely exchange between the family members and the medical staff. It may have also reduced the isolation of the patient participants and their family members. Even though user satisfaction with LEAN remained high throughout the program, it dropped considerably from 98.4% (62/63) to 67.4% (66/98) for patients and 100% (77/77) to 80.0% (100/125) for the lay health supporters from phase 1 to phase 3 (Multimedia Appendix 6, Table S2), In future studies, the frequency, content, and timing of texting should be optimized for each set to reduce user fatigue [52]. Finally, considering the reduced cognition of some people with schizophrenia, we used the most rudimentary model of text messaging without individual tailoring and smartphone-based apps. Other settings should fully consider the cost, feasibility, and acceptability in their community and family context to determine the optimal program mode.

Conclusions

In our study, we found that our intervention featuring mobile texting messages and lay health workers in a resource-poor community setting was more effective in improving medication adherence, symptoms, and rehospitalization than the 686 Program alone in long-term implementation. The experiences of LEAN can potentially be widely applicable to improve medication adherence of other chronic diseases in other resource-poor settings.

Acknowledgments

The project received grant support from the National Institute on Aging of the National Institutes of Health (grant number: P30AG047845; principal investigator: DX) and the China Medical Board (grant number: 12-114; principal investigator: WG). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

We thank the all the participants, the Liuyang Mental Health Center (Meng Dai, Li Chen) for providing patient assessment and policy support, mental health administrators from 9 township health centers (Zhong Huang, Huakun Li, Yanjiang Li, Xianyong Li, Jiaona Liu, Change Lu, Hao Luo, Xiao Pan, Huiren Shao, Shuyi Tang, Baojing Tian, Xiang Wang, Qing Zeng) for coordinating and assisting in our fieldwork and home visits, the short message development and management group (Yongsheng Tong, Tianlin Ma, Junren Wang, Yun Luo, Jinan Long, Jiali Ye, Zihui Liang), and the LEAN project volunteer group (Chao Zhang, Min Yu, Qiong Tang, Yushi Mo, Anwen Yang, Bingwei Tang, Xin Jin, Yue Zeng, Weiyue Long, Huilan Tang, Wei Liu, Tingting Li, Xiaoyan Guo, Cheng Hu, Xing Liu, Nan Xia, Jingyan Fan, Jingjing Yao, Ziwei Ye, Hao Fang, Weina Zhang, Fan Yang, Wenqing Xue, Lu Tang, Fengying Bi, Qian Zhou, Xihong Wu, Xingkui Pan, Jian Xu, Yuhan Pi, Shilin Yin, Bo Chen, Donglan Zhang) for field data collection and management.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The treatment of missing data.
[DOCX File , 413 KB - jmir_v22i12e22631_app1.docx]

Multimedia Appendix 2
Frequency of participants with different pill-count adherence in phase 1 and phase 3.
[DOCX File , 55 KB - jmir_v22i12e22631_app2.docx]

Multimedia Appendix 3
Reasons for discontinuing.
[DOCX File , 16 KB - jmir_v22i12e22631_app3.docx]

Multimedia Appendix 4
Relationship between WHODAS interviewees and patients.
[DOCX File , 16 KB - jmir_v22i12e22631_app4.docx]

Multimedia Appendix 5
Sensitivity analyses.
[DOCX File , 21 KB - jmir_v22i12e22631_app5.docx]

Multimedia Appendix 6
Process indicators.
[DOCX File , 31 KB - jmir_v22i12e22631_app6.docx]

Multimedia Appendix 7
Cohen d effect size of the primary outcome in other research.
[DOCX File , 16 KB - jmir_v22i12e22631_app7.docx]

References

1. McGrath J, Saha S, Chant D, Welham J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiol Rev 2008;30:67-76. [doi: 10.1093/epirev/mxn001] [Medline: 18480098]
2. Leucht S, Cipriani A, Spineli L, Mavridis D, Örey D, Richter F, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. The Lancet 2013 Sep;382(9896):951-962. [doi: 10.1016/S0140-6736(13)60733-3]
3. Iasevoli F, Giordano S, Balletta R, Latte G, Formato MV, Prinzivalli E, et al. Treatment resistant schizophrenia is associated with the worst community functioning among severely-ill highly-disabling psychiatric conditions and is the most relevant predictor of poorer achievements in functional milestones. Prog Neuropsychopharmacol Biol Psychiatry 2016 Feb 04;65:34-48. [doi: 10.1016/j.pnpbp.2015.08.010] [Medline: 26320028]
4. Goff DC, Hill M, Freudenreich O. Strategies for improving treatment adherence in schizophrenia and schizoaffective disorder. J Clin Psychiatry 2010;71 Suppl 2:20-26 [FREE Full text] [doi: 10.4088/JCP.9096su1cc.04] [Medline: 21190649]
5. mhGAP Mental Health Gap Action Programme—scaling up care for mental, neurological and substance use disorders. World Health Organization. URL: https://www.who.int/mental_health/mhgap_final_english.pdf?ua=1 [accessed 2020-09-23]
6. Good BJ, Good MD. Significance of the 686 Program for China and for global mental health. Shanghai Arch Psychiatry 2012 Jun;24(3):175-177 [FREE Full text] [doi: 10.3969/j.issn.1002-0829.2012.03.008] [Medline: 25324623]
7. Kang HM, Guan LL, Wang X, Wu XM. Economic condition， basic medical insurance and medication among patients with severe mental illness in 31 demonstration cities for“686 program”implementation [In Chinese]. Chin J Public Health 2015;31:997-1001. [doi: 10.11847/zgggws2015-31-08-06]
8. Wu XM, Ma M, Wang X, Zhang WF. Management and services for psychosis in the People?S Republic of China in 2017[In Chinese]. China Journal Psychiatry 2019;52:82-88. [doi: 10.3760/cma.j.issn.1006-7884.2019.01.008]
9. Liang D, Mays VM, Hwang W. Integrated mental health services in China: challenges and planning for the future. Health Policy Plan 2018 Jan 01;33(1):107-122 [FREE Full text] [doi: 10.1093/heapol/czx137] [Medline: 29040516]
10. Pijnenborg GHM, Withaar FK, Brouwer WH, Timmerman ME, van DBRJ, Evans JJ. The efficacy of SMS text messages to compensate for the effects of cognitive impairments in schizophrenia. Br J Clin Psychol 2010 Jun;49(Pt 2):259-274. [doi: 10.1348/014466509X467828] [Medline: 19735607]

11. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012 May;38(3):414-425 [FREE Full text] [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
12. Montes JM, Medina E, Gomez-Beneyto M, Maurino J. A short message service (SMS)-based strategy for enhancing adherence to antipsychotic medication in schizophrenia. Psychiatry Res 2012 Dec 30;200(2-3):89-95. [doi: 10.1016/j.psychres.2012.07.034] [Medline: 22901437]
13. Španiel F, Hrdlička J, Novák T, Kožený J, Höschl C, Mohr P, et al. Effectiveness of the information technology-aided program of relapse prevention in schizophrenia (ITAREPS): a randomized, controlled, double-blind study. J Psychiatr Pract 2012 Jul;18(4):269-280. [doi: 10.1097/01.pra.0000416017.45591.c1] [Medline: 22805901]
14. Beebe L, Smith KD, Phillips C. A comparison of telephone and texting interventions for persons with schizophrenia spectrum disorders. Issues Ment Health Nurs 2014 May;35(5):323-329. [doi: 10.3109/01612840.2013.863412] [Medline: 24766166]
15. Wiersma D, Wanderling J, Dragomirecka E, Ganev K, Harrison G, An Der Heiden W, et al. Social disability in schizophrenia: its development and prediction over 15 years in incidence cohorts in six European centres. Psychol Med 2000 Sep;30(5):1155-1167. [doi: 10.1017/s0033291799002627] [Medline: 12027051]
16. Xu DR, Gong W, Caine ED, Xiao S, Hughes JP, Ng M, et al. Lay health supporters aided by a mobile phone messaging system to improve care of villagers with schizophrenia in Liuyang, China: protocol for a randomised control trial. BMJ Open 2016 Jan 20;6(1):e010120 [FREE Full text] [doi: 10.1136/bmjopen-2015-010120] [Medline: 26792221]
17. Xu DR, Xiao S, He H, Caine ED, Gloyd S, Simoni J, et al. Lay health supporters aided by mobile text messaging to improve adherence, symptoms, and functioning among people with schizophrenia in a resource-poor community in rural China (LEAN): A randomized controlled trial. PLoS Med 2019 Apr 23;16(4):e1002785. [doi: 10.1371/journal.pmed.1002785]
18. Beard E, Lewis JJ, Copas A, Davey C, Osrin D, Baio G, et al. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. Trials 2015;16:353 [FREE Full text] [doi: 10.1186/s13063-015-0839-2] [Medline: 26278881]
19. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organization. URL: https://www.who.int/classifications/icd/en/bluebook.pdf [accessed 2020-11-22]
20. Becker MH, Maiman LA. Sociobehavioral determinants of compliance with health and medical care recommendations. Med Care 1975 Jan;13(1):10-24. [doi: 10.1097/00005650-197501000-00002] [Medline: 1089182]
21. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
22. Finfgeld DL, Wongvatunyu S, Conn VS, Grando VT, Russell CL. Health belief model and reversal theory: a comparative analysis. J Adv Nurs 2003 Aug;43(3):288-297. [doi: 10.1046/j.1365-2648.2003.02712.x] [Medline: 12859788]
23. Zhao M, Gong WJ, Xu DR, Huang ZP, Hu B. Nie J The Training Effecton Text-messaging Skills of Patients with Schizophreniain rural China and Its Influencing Factors In Chinese. China Journal of Health Psychology 2018;26(10):1450-1454. [doi: 10.13342/j.cnki.cjhp.2018.10.003]
24. Byerly MJ, Nakonezny PA, Rush AJ. The Brief Adherence Rating Scale (BARS) validated against electronic monitoring in assessing the antipsychotic medication adherence of outpatients with schizophrenia and schizoaffective disorder. Schizophr Res 2008 Mar;100(1-3):60-69 [FREE Full text] [doi: 10.1016/j.schres.2007.12.470] [Medline: 18255269]
25. Cheng HL, Yu YW. [Validation of the Chinese version of "the Drug Attitude Inventory"]. Kaohsiung J Med Sci 1997 Jun;13(6):370-377. [Medline: 9251303]
26. Stjernswärd S, Persson K, Nielsen R, Tuninger E, Levander S. A modified Drug Attitude Inventory used in long-term patients in sheltered housing. Eur Neuropsychopharmacol 2013 Oct;23(10):1296-1299. [doi: 10.1016/j.euroneuro.2012.11.011] [Medline: 23265955]
27. Guilera G, Gómez-Benito J, Pino O, Rojo JE, Cuesta MJ, Martínez-Arán A, et al. Utility of the World Health Organization Disability Assessment Schedule II in schizophrenia. Schizophr Res 2012 Jul;138(2-3):240-247. [doi: 10.1016/j.schres.2012.03.031] [Medline: 22521724]
28. Sjonnesen K, Bulloch AGM, Williams J, Lavorato D, B Patten S. Characterization of disability in Canadians with mental disorders using an abbreviated version of a DSM-5 emerging measure: the 12-Item WHO Disability Assessment Schedule (WHODAS) 2.0. Can J Psychiatry 2016 Apr;61(4):227-235 [FREE Full text] [doi: 10.1177/0706743716632514] [Medline: 27254415]
29. Haro JM, Kamath SA, Ochoa S, Novick D, Rele K, Fargas A, SOHO Study Group. The Clinical Global Impression-Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. Acta Psychiatr Scand Suppl 2003(416):16-23. [doi: 10.1034/j.1600-0447.107.s416.5.x] [Medline: 12755850]
30. Fenton WS, Blyler CR, Heinssen RK. Determinants of medication compliance in schizophrenia: empirical and clinical findings. Schizophr Bull 1997 Jan 01;23(4):637-651. [doi: 10.1093/schbul/23.4.637] [Medline: 9366000]
31. Waddell L, Taylor M. A new self-rating scale for detecting atypical or second-generation antipsychotic side effects. J Psychopharmacol 2008 May;22(3):238-243. [doi: 10.1177/0269881107087976] [Medline: 18541624]
32. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 2015 Feb 06;350:h391 [FREE Full text] [Medline: 25662947]

33. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007 Feb;28(2):182-191. [doi: 10.1016/j.cct.2006.05.007] [Medline: 16829207]
34. Barker D, McElduff P, D'Este C, Campbell MJ. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. BMC Med Res Methodol 2016 Jun 06;16(1):69 [FREE Full text] [doi: 10.1186/s12874-016-0176-5] [Medline: 27267471]
35. Khadka A, Perales NA, Wei DJ, Gage AD, Haber N, Verguet S, et al. Malaria control across borders: quasi-experimental evidence from the Trans-Kunene malaria initiative (TKMI). Malar J 2018 Jun 04;17(1):224 [FREE Full text] [doi: 10.1186/s12936-018-2368-4] [Medline: 29866113]
36. Li F, Turner EL, Preisser JS. Sample size determination for GEE analyses of stepped wedge cluster randomized trials. Biometrics 2018 Dec 19;74(4):1450-1458 [FREE Full text] [doi: 10.1111/biom.12918] [Medline: 29921006]
37. Hollis S, Campbell F. What is meant by intention to treat analysis? survey of published randomised controlled trials. BMJ 1999 Sep 11;319(7211):670-674 [FREE Full text] [doi: 10.1136/bmj.319.7211.670] [Medline: 10480822]
38. van Buuren S. Flexible Imputation of Missing Data. Boca Raton, Florida: CRC Press; 2012.
39. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Lawrence Earlbaum Associates; 1988.
40. Välimäki M, Kannisto KA, Vahlberg T, Hätönen H, Adams CE. Short text messages to encourage adherence to medication and follow-up for people with psychosis (Mobile.Net): randomized controlled trial in Finland. J Med Internet Res 2017 Jul 12;19(7):e245 [FREE Full text] [doi: 10.2196/jmir.7028] [Medline: 28701292]
41. Chen R, Liou T, Miao N, Chang K, Yen C, Liao H, et al. Using World Health Organization Disability Assessment Schedule 2.0 in people with schizophrenia: a 4-year follow-up. Eur Arch Psychiatry Clin Neurosci 2020 Apr 22;270(3):301-310. [doi: 10.1007/s00406-019-01000-5] [Medline: 31011812]
42. Dan A, Kumar S, Avasthi A, Grover S. A comparative study on quality of life of patients of schizophrenia with and without depression. Psychiatry Res 2011 Sep 30;189(2):185-189. [doi: 10.1016/j.psychres.2011.02.017] [Medline: 21453977]
43. Matza LS, Phillips GA, Revicki DA, Ascher-Svanum H, Kaiser D, Stauffer V, et al. Development of a clinician questionnaire and patient interview to assess reasons for antipsychotic discontinuation. Psychiatry Res 2011 Oct 30;189(3):463-468. [doi: 10.1016/j.psychres.2011.05.030] [Medline: 21684017]
44. Matza L, Phillips, Revicki, Ascher-Svanum H, Malley, Palsgrove, Faries, Stauffer, Awad, Keefe, Naber. Validation of a patient interview for assessing reasons for antipsychotic discontinuation and continuation. Patient Prefer Adherence 2012;6:521-532 [FREE Full text] [doi: 10.2147/PPA.S25635] [Medline: 22879738]
45. Matza LS, Phillips GA, Revicki DA, Ascher-Svanum H, Malley KG, Palsgrove AC, et al. Validation of a clinician questionnaire to assess reasons for antipsychotic discontinuation and continuation among patients with schizophrenia. Psychiatry Res 2012 Dec 30;200(2-3):835-842. [doi: 10.1016/j.psychres.2012.05.005] [Medline: 22841345]
46. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
47. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). Clin Trials 2014 Jun 5;11(4):426-434. [doi: 10.1177/1740774514536795] [Medline: 24902922]
48. Xu DR, Gong W, Gloyd S, Caine ED, Simoni J, Hughes JP, et al. Measuring adherence to antipsychotic medications for schizophrenia: concordance and validity among a community sample in rural China. Schizophr Res 2018 Nov;201:307-314 [FREE Full text] [doi: 10.1016/j.schres.2018.05.014] [Medline: 29807806]
49. Yang L, Li M, Tao L, Zhang M, Nicholl MD, Dong P. Cost-effectiveness of long-acting risperidone injection versus alternative atypical antipsychotic agents in patients with schizophrenia in China. Value Health 2009;12 Suppl 3:S66-S69 [FREE Full text] [doi: 10.1111/j.1524-4733.2009.00630.x] [Medline: 20586985]
50. Leucht C, Heres S, Kane JM, Kissling W, Davis JM, Leucht S. Oral versus depot antipsychotic drugs for schizophrenia--a critical systematic review and meta-analysis of randomised long-term trials. Schizophr Res 2011 Apr;127(1-3):83-92. [doi: 10.1016/j.schres.2010.11.020] [Medline: 21257294]
51. Xiang Y, Weng Y, Leung C, Tang W, Ungvari G. Clinical and social correlates with the use of depot antipsychotic drugs in outpatients with schizophrenia in China. Int J Clin Pharmacol Ther 2008 May;46(5):245-251. [doi: 10.5414/cpp46245] [Medline: 18538110]
52. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]

Abbreviations

BARS: Brief Adherence Rating Scale
CGI: Clinical Global Impression
DAI-10: Drug Attitude Inventory-10
DAS: Disability Assessment Schedule

SMS: short message service
WHO: World Health Organization

Edited by G Eysenbach, R Kukafka; submitted 18.07.20; peer-reviewed by Y Tong, E Moran; comments to author 10.08.20; revised version received 23.09.20; accepted 29.10.20; published 01.12.20.

Please cite as:
Cai Y, Gong W, He H, Hughes JP, Simoni J, Xiao S, Gloyd S, Lin M, Deng X, Liang Z, He W, Dai B, Liao J, Hao Y, Xu D(
Mobile Texting and Lay Health Supporters to Improve Schizophrenia Care in a Resource-Poor Community in Rural China (LEAN Trial): Randomized Controlled Trial Extended Implementation
J Med Internet Res 2020;22(12):e22631
URL: https://www.jmir.org/2020/12/e22631
doi:10.2196/22631
PMID:33258788

Original Paper

Stakeholder Feedback of Electronic Medication Adherence Products: Qualitative Analysis

Sadaf Faisal[1], BPharm; Jessica Ivo[1], BSc; Aidan McDougall[2], BSc; Tejal Patel[1,2,3], PharmD

[1]School of Pharmacy, University of Waterloo, Kitchener, ON, Canada
[2]Centre for Family Medicine Family Health Team, Kitchener, ON, Canada
[3]Schlegel-University of Waterloo Research Institute of Aging, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Tejal Patel, PharmD
School of Pharmacy
University of Waterloo
10 Victoria St S
Kitchener, ON, N2G 1C5
Canada
Phone: 1 519 888 4567 ext 21337
Email: t5patel@uwaterloo.ca

Abstract

Background: Medication management among older adults continues to be a challenge, and innovative electronic medication adherence products have been developed to address this need.

Objective: The aim of this study is to examine user experience with electronic medication adherence products, with particular emphasis on features, usefulness, and preferences.

Methods: Older adults, caregivers, and health care providers tested the usability of 22 electronic medication adherence products. After testing 5 products, participants were invited to participate in a one-on-one interview to investigate their perceptions and experiences with the features, usefulness, and preference for electronic medication adherence products tested. The interviews were audio recorded, transcribed, and analyzed using exploratory inductive coding to generate themes. The first 13 interviews were independently coded by 2 researchers. The percentage agreement and Cohen kappa after analyzing those interviews were 79% and 0.79, respectively. A single researcher analyzed the remaining interviews.

Results: Of the 37 participants, 21 (57%) were older adults, 5 (14%) were caregivers, and 11 (30%) were health care providers. The themes and subthemes generated from the qualitative analysis included product factors (subthemes: simplicity and product features, including availability and usability of alarms, portability, restricted access to medications, and storage capacity) and user factors (subthemes: sentiment, affordability, physical and cognitive capability, and technology literacy and learnability).

Conclusions: Electronic medication adherence products have the potential to enable independent medication management in older adults. The choice of a particular product should be made after considering individual preferences for product features, affordability, and the sentiment of the users. Older adults, caregivers, and health care providers prefer electronic medication adherence products that are simple to set up and use, are portable, have easy-to-access medication compartments, are secure, and have adequate storage capacity.

(J Med Internet Res 2020;22(12):e18074) doi:10.2196/18074

KEYWORDS

medication nonadherence; technology; aged; patient preferences; eHealth; qualitative research; adherence

Introduction

Background

The older adult population is increasing rapidly worldwide, with a projection of approximately 1.5 billion individuals aged 65 years or older by 2050 [1]. In North America and Europe alone, individuals aged 65 years or older will account for 19.1% of the total population by 2050 [1]. In 2016, a Canadian study showed that 65.7% of adults aged 65 years or older were prescribed 5 or more medications, 26.5% were prescribed 10 or more medications, and 8.4% were prescribed 15 or more medications [2]. Approximately 50% of older adults are nonadherent to their medications [3]. Poor medication adherence leads to nonoptimal management of diseases; increased emergency room visits; hospitalization; and, ultimately, poor quality of life [4].

The incidence of nonadherence is highly prevalent in the geriatric population because of multiple factors [5]. As people age, they are diagnosed with multiple chronic health conditions that often require complex medication regimens, with multiple medications, various dosage forms, and complicated medication schedules [6]. In addition, older adults face significant issues with medication management because of impaired functional capabilities, such as poor vision, hearing loss, dexterity issues, or cognitive capabilities [6-8]. These functional and cognitive capabilities are imperative for managing complex medication regimens; therefore, older adults with such deficits are at higher risk of medication mismanagement [9].

Numerous medication management strategies are being used by older adults to improve adherence, such as pillboxes, dosettes, blister packaging provided by pharmacies, reminders or alarms, and automated pill dispensers [10]. Electronic medication adherence products are electronic products that are capable of dispensing medications in an attempt to address the limitations related to medication management capacity. These products have a multitude of features, including integrated alarms, multiple compartments, automatic components, security features, adherence report–generating features, and cloud connectivity [11]. However, there is limited research investigating the impact of electronic medication adherence products on medication adherence and, in particular, their use in older adults. In a previous qualitative study evaluating the views of patients and health care professionals with regard to electronic multicompartment medication devices, participants were asked to interact with 8 devices for 5 to 10 min using the think-aloud method to promote discussion in a focus group or in one-on-one interviews [12]. However, this study did not test whether participants were able to use the device for its intended purpose, for example, whether the participant could accurately fill the device, set the alarm, and access the medications from the device. Without testing these particular features, users may have difficulty fully establishing the usability and workload of a device. Therefore, we designed a mixed methods study to assess the usability, workload, and user experience of electronic medication adherence products for 3 stakeholder groups: older adults, caregivers, and health care providers. Caregivers and health care providers were included in conjunction with older adults to provide a holistic understanding of product use. Caregivers play a significant role in medication management of older adults and may at times be responsible for organizing medication administration aids, helping to administer medications, or many other medication management activities [13]. Similarly, health care providers often recommend medication adherence aids to patients experiencing nonadherence. The results of the usability and workload analysis from this mixed methods study have been published elsewhere [14].

Objectives

The aim of this study is to explore the feedback and experience of stakeholders with regard to electronic medication adherence products, with particular emphasis on features, usefulness, and preferences.

Methods

Study Design

We conducted an exploratory qualitative study using semistructured interviews of stakeholders to assess the usefulness of, preference for, and features of electronic medication adherence products as part of a larger mixed methods study.

Ethical Consideration

This study was reviewed by the Office of Research and Ethics of the University of Waterloo, and the study received ethics approval before recruitment and data collection. All participants were informed of the study, and they provided written consent before enrolling in the study.

Study Participants

Purposive sampling techniques were used to recruit 3 types of stakeholders: older adults, caregivers, and health care providers. Individuals were eligible to take part in the study if they were (1) an adult aged 65 years or older who was taking one or more medications regularly and living independently, known as an *older adult participant*; (2) an adult who may or may not be living with the older adult participant and was involved in their medication management, known as a *caregiver*; or (3) health care providers (eg, physicians, pharmacists, nurses, and/or occupational therapists) who are involved in caring for the geriatric patient population in their practice and who may or may not have been involved in the direct care of an older adult participant. In addition, eligible participants were required to be able to speak English. Recruitment flyers and emails to organizations were used to recruit older adults and caregivers. In addition, a participant pool of older adults who indicated interest in participating in this study was approached via telephone. Health care providers were approached via email.

Study Procedure

A total of 23 electronic medication adherence products were purchased; however, 1 was nonfunctional, and 1 was found to be nonelectronic but was still tested. Thus, 22 electronic medication adherence products were tested by participants (Table 1 provides a list of the electronic medication adherence products used). Products were chosen for purchase with the aim of representing different features, such as the number and type of compartments, audio and visual reminders, and medication tracking. Participants were invited to attend a 3-hour testing session at the School of Pharmacy Research Laboratory of the University of Waterloo. At the start of product testing, participants were introduced to a mock medication regimen containing placebo tablets, candy (Tic Tac), and placebo capsules and asked to familiarize themselves with the regimen by reading the labels of the medications. The mock medication regimen was designed by a pharmacist researcher and 2 research assistants after reviewing the literature on the most commonly dispensed medications for older adults and by examining the average number of medications taken by this population. Each participant tested 5 randomly selected products sequentially. At the start of each product test, participants were provided with instruction manuals made by the manufacturer. If the manufacturer's instructions were not enclosed with the product, researchers found and printed web-based instructions for the participant. The testing tasks varied for each product but included steps such as opening and filling the medication tray, setting up the alarm, locking the device, and accessing the placebo pills from the device. Participants were given as much time as they needed to address all testing tasks and were not provided assistance unless requested.

Table 1. Electronic medication adherence products used in this study.

Name	Number of compartments[a]	Locking feature[b]	Portability[c]	Price, US $
GMS[d] Med-e-lert automatic pill dispenser	28	Yes	No	70-109
LiveFine automatic pill dispenser and reminder	28	Yes	No	70-109
MedReady 1700 automated medication dispenser	28	Yes	No	≥109
MedSmart med-reminder and dispensing system	29	Yes	Yes	≥109
e-pill MedTime Station automatic pill dispenser with tipper	28	Yes	No	≥109
TimerCap travel size	1[e]	No	Yes	<30
TimerCap universal size	1[e]	No	Yes	<30
Jones medication adherence system 14-unit card	14	No	Yes	N/A[f]
Reizen vibrating pill box	5	No	Yes	<30
VitaCarry advanced pill case	7	No	Yes	30-69
Nishiki round pill box with alarm	7	No	Yes	<30
MedGlider system 1 with talking reminder	4	No	Yes	30-69
Patterson medical tabtime super 8	8	No	Yes	30-69
100-Hour pill reminder	3	No	Yes	<30
MedQ smart pill box	14	No	Yes	70-109
e-pill MedGlider home medication management system	7	No	Yes	30-69
MedCentre system	30	No	Yes	30-69
eNNOVEA Weekly Planner With Advanced Auto Reminder	14	No	Yes	70-109
e-pill Multi-alarm pocket XL	7	No	Yes	30-69
6 Grid pill storage case with alarm	6	No	Yes	<30
Itzbeen pocket doctor	0	No	Yes	<30
e-pill Weekly Accutab pill box organizer system[g]	21	No	No	30-69

[a] The total number of compartments a product contains to store medications.

[b] If the product has the ability to lock users from accessing their medications.

[c] If the product can be carried outside of the house using a purse or small bag.

[d] GMS: Group Medical Supply.

[e] Device has one compartment that can be accessed multiple times.

[f] N/A: not applicable.

[g] Device was found to be nonelectronic.

Following this testing phase, participants were invited to an optional one-on-one semistructured interview. The interview questions were developed by the research team and finalized through agreement. The research team consisted of 3 pharmacists and 1 occupational therapist who works closely with the geriatric patient population, 1 PhD student with extensive experience as a practicing pharmacist, 1 research assistant with 2 years of research experience, and 1 co-operative education student. All team members were female. Interview questions were developed to elicit information and feedback related to the experience of participants using the products (Textbox 1 provides a complete interview guide).

Textbox 1. Interview guide.

Question 1

- What was one feature of any of the product that you liked the most? Why did you like this feature?
- What was one feature of any of the product that you did not like or liked the least? Why did you not like this feature?

Question 2

- What feature on any product did you find was the most useful? Why did you find this the most useful?
- What feature on any product did you find was the least useful? Why did you find this the least useful?

Question 3

- Do you use a medication administration aid currently? What do you use, and why do you use it?

Question 4

- Which of the products that you tested would you use for your own medication regimen? Why or why not?
- OR, would you recommend any of these products to your patients or parents to use for their medication regimen? Why or why not?

Question 5

- If you use an aid to help with your own medications, please compare this with the ones you tested today. Which do you prefer and why?

Data Collection

A total of 2 research team members (AM and JI) conducted the interviews. All interviews took place between August 2018 and December 2018. If an older adult–caregiver dyad participated in the study, each person was interviewed separately. Each interview lasted for 15 to 30 min. All interviews were audio recorded using a Sony IC recorder ICD-PX470 and transcribed verbatim by 1 of 2 research team members (JI or AM), after which the accuracy of the transcription was established by the other research member. The interviews were transcribed using Microsoft Office Word.

Data Analysis

The 6-phase framework by Braun et al [15] was used to perform thematic analysis [16]. The first interview transcript was individually coded by 2 research team members (JI and SF), after which preliminary codes were identified and compared with ensure coding consistency. The next 13 interviews were independently coded by the 2 researchers. A codebook was created to finalize the list of codes, and it contained the code name, code description, and quotes from the interview data. Percentage agreement and Cohen kappa were calculated between coders to ensure consistency in the first 13 interviews and was found to be 79% and 0.79, respectively. Any disagreement was discussed between the 2 researchers and resolved. The plan to bring any unresolved disagreements to a third researcher (TP) was not required. The remaining interviews were then analyzed by a single researcher (SF).

Results

Participant Demographics

A total of 39 participants were recruited for the larger study, of which 2 (5%) participants refused to participate in the optional one-on-one interview. Participants were not asked for a reason as to why they chose not to participate in the interview. Of the 37 participants, 21 (57%) were older adults, 5 (14%) were caregivers, and 11 (30%) were health care providers. There were 3 older adult–caregiver dyads. These dyads did not test the same products; they tested products independently and were interviewed individually (Table 2 shows the study participant characteristics).

Table 2. Study participant characteristics (N=37).

Participant characteristics	Values
Older adults (n=21)	
Gender, n (%)	
Male	10 (48)
Female	11 (52)
Age (years)	
Mean (SD)	75 (6.5)
Range	65-87
Total number of medications taken per participant	
Mean (SD)	7.7 (3.3)
Range	1-13
Number of medications taken per participant by their type, n (%)	
≥5 Rx[a], supplement, OTC[b], and herbal	17 (81)
≥5 Rx	10 (48)
≥5 Supplement, OTC, and herbal	6 (29)
Medication schedule, n (%)	
Once daily	4 (19)
More than once daily	17 (81)
Medication aids use, n (%)	
Yes	15 (71)
No	6 (29)
Medication aids used in combination, n (%)	
Yes	4 (19)
Caregivers (n=5)	
Gender reported by the caregiver for their patient, n (%)	
Male	4 (80)
Female	1 (20)
Age reported by the caregiver for their patient (years)	
Mean (SD)	73 (4.49)
Range	69-79
Caregiver's relationship with the patient, n (%)	
Family member	5 (100)
Friend	0 (0)
Other	0 (0)
Health care providers (n=11)	
Gender, n (%)	
Male	2 (18)
Female	9 (82)
Occupation, n (%)	
Pharmacist	8 (73)
Pharmacy student	1 (9)
Occupational therapist	2 (18)
Years of practice	

Participant characteristics	Values
Mean (SD)	8.8 (10.5)
Range	0[c]-37
Number of older adults worked with in a typical week, n (%)	
<10	1 (9)
10-20	4 (36)
20-30	1 (9)
>30	5 (45)
Medication aids recommendation, n (%)	
Yes	11 (100)

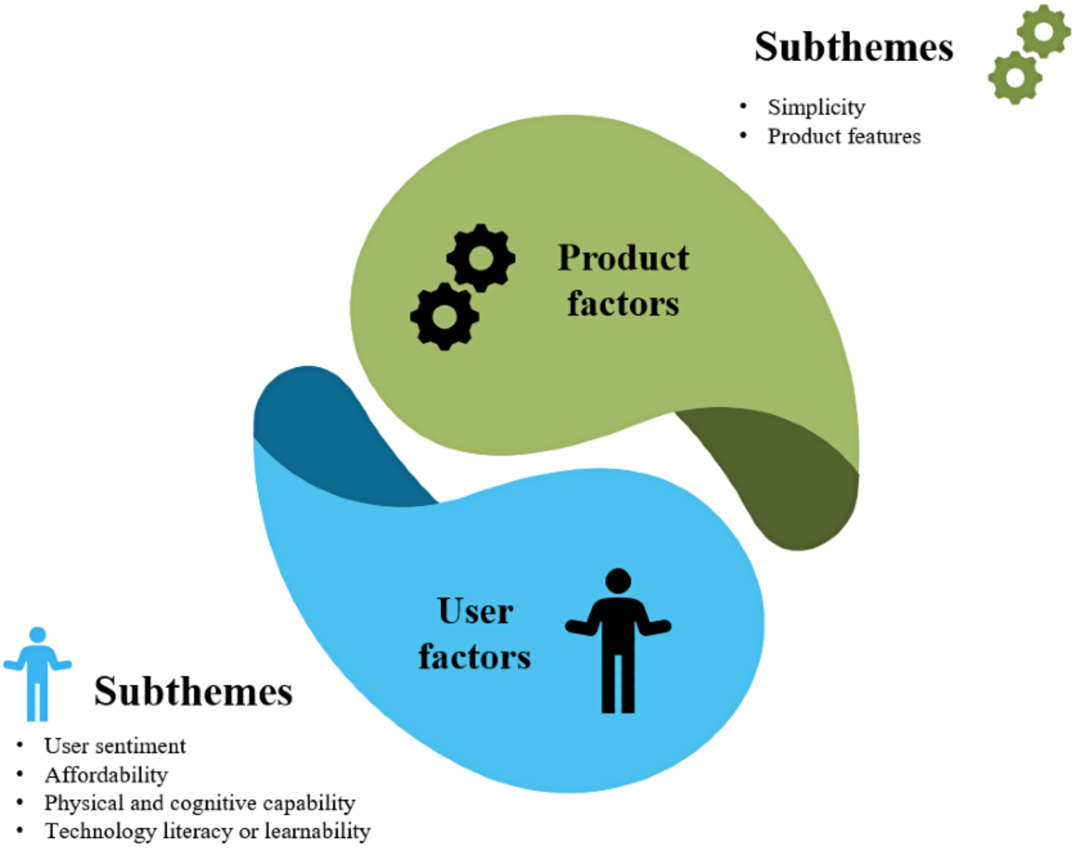
[a] Rx: prescription medications.

[b] OTC: over the counter.

[c] One health care provider was a pharmacy student and, thus, had 0 years of practice as a registered pharmacist.

A total of 39 codes were identified after the analysis. Codes were grouped into sub-themes, from which 2 themes and 6 sub-themes emerged (Figure 1 shows the themes and subthemes).

Figure 1. Themes and subthemes related to user experience with electronic adherence products.



Theme 1: Product Factors

Product factors was 1 of the 2 themes that emerged from our interview analysis. Product factors are factors associated with a product that can impact its use. The theme product factors were further divided into 2 subthemes: simplicity and product features.

Simplicity

Some electronic medication adherence products required multiple steps to set up, and although detailed information was provided by the manufacturer, it was not always perceived to be helpful by stakeholders. Stakeholders found some products easy to set up and simple to use, whereas they found other products to be complex. In general, the initial impression of complexity or simplicity impacted the overall impression and projected use of the product. Stakeholders did not appreciate products that required repetitive reviews of instructions to set up. One participant expressed their concerns through the following quote:

> *[This product is] just fiddley and difficult. I had to go over the instructions several times to understand that you had to lock, it wasn't simple. And people, older people especially if they have a... long term illness, [they] want things that are simple. They don't want things that are going to be fiddley or [a] pain... [or] annoying to operate.* [Older adult 003]

Product Features

Product features are the characteristics, components, or capabilities of an electronic medication adherence product. This subtheme included the availability and usability of alarms, portability, restricted access to medication, and storage capacity.

Availability and Usability of Alarms

Stakeholders enjoyed electronic medication adherence products with audio or visual reminders, as they helped improve adherence. One participant described the importance of having the alarm as follows:

> *But no alarm, and that's the downside to this [electronic medication adherence product]. That there's no alarm on it. So, often I forget to take my Levodopa that I should be taking.* [Older adult 004]

However, the initial setup required for the alarm was considered to be an integral factor when choosing a product. Some products had small buttons or complex instructions that were not appealing to stakeholders.

Portability

In general, study participants commented negatively on products that were not portable, stating that products were too large, too noticeable, or visible in their home environment. Having a portable electronic medication adherence product was described as an important feature to consider when choosing a product for medication management, as a nonportable product may lead to missing a dose:

> *Like I'm going out for lunch with my women's group. I can't take this flying saucer with me. It's not going to fit in my...the purse that I want to bring. So, I'm gonna miss... I'm probably gonna miss the dose by being out.* [Older adult 027]

Restricted Access to Medication

Another subtheme that emerged was the ability of a product to provide restricted access to medications, in turn allowing for secure medication administration and improved safety, that is, decrease the risk of inadvertently taking the wrong medication at the wrong time or double dosing. The ability to dispense the right dose at the right time was mentioned and preferred by most of our participants, as demonstrated by the following quote:

> *I liked that you couldn't open it if you, like you...it's impossible to take a second dose at the wrong time. It's the only one of them that makes it so that you can't essentially overdose on something.* [Older adult 028]

Storage Capacity

In this study, older adult participants took an average of 8 medications per day (range 1-13). Therefore, it is unsurprising that one feature frequently mentioned by participants when considering an electronic medication adherence product was its storage capacity. Storage capacity includes the number of the compartments and the number of days a product can store medications for. One caregiver discussed how none of the 5 products they tested during the study were suitable for the medication routine of their care recipient:

> *I couldn't use any of them for [my husband] because nothing [allows for] pills 8 times during the day. [For] the pillbox that we do use, I had to buy two sets. I have two different sets at home [to] accommodate all the pills [and] different times.* [Caregiver 008]

Health care providers similarly viewed a product's storage capacity as important. One health care provider was quite pleased with how accommodating a product was with regard to its flexibility in compartments. This particular product had 28 compartments and allowed for 6 daily alarms:

> *Most of the patients who are senior, they are usually on 3-6 medications. So, and on multiple dosing, so 2 times a day, 3 times a day, 4 times a day. So, this [product] when you are [done with the] set up, it's very well laid out for them.* [Pharmacist 023]

Overall, participants preferred electronic medication adherence products that had the ability to store multiple medications for several days or weeks. Users often found this feature very convenient and preferable when choosing a medication management aid.

Theme 2: User Factors

User factors are defined as factors or abilities that can impact how a user uses the product effectively. User factors comprised 4 subthemes, including sentiment, affordability, physical and cognitive capability, and technology literacy or learnability.

Sentiment

Our interview analysis identified that participants felt various sentiments when using electronic medication adherence products. Sentiments can be defined as "an attitude, thought or judgment prompted by feeling" [17]. Some participants felt a sense of safety or peace of mind, whereas others were worried about the privacy of their medication intake process. A few participants also felt frustrated because the product was very complicated to set up and use, which impacted their overall impression of the product. These feelings were grouped under the following subthemes: sense of assurance, privacy, and frustration.

Sense of Assurance

An important aspect mentioned by participants when choosing an electronic medication adherence product was having a sense of assurance, so users did not have to worry about missing their medication dose. This assurance also provided participants with a sense of safety. For instance, one participant stated:

> *You don’t have to worry about “did I take my meds.”* [Older adult 001]

Health care providers shared the same feeling about having peace of mind when recommending electronic medication adherence products for their patients. One health care provider stated:

> *[The product] lets you know when you last took [the dose]... Even though you have an alarm telling you to take it... some people might forget that they actually took it and then might go to take... or check to see if there is another dose. So, this helps to know that it’s only been two or four or six hours since they last took it.* [Pharmacist 006]

This feeling of assurance was also associated with familiarity of a certain feature within an electronic medication adherence product. One participant preferred products that were similar to the pillbox they used at home and felt very comfortable using that particular product:

> *I like the pillbox [product name] because it’s huge and I'm used to doing that.* [Older adult 002]

Privacy

During this study, we tested products that ranged from small pillboxes to large automated dispensers. Most of our study participants disliked the idea of using large products and preferred products that were less noticeable. Participants found that large products brought attention to their medication regimen, a process that they would like to keep private. An example of a quote expressing concern about privacy is given below:

> *It was unobtrusive it wouldn't take up a lot of space to put it on the counter and nobody would really notice it. So, I like that about it.* [Older adult 003]

User Frustration

Another subtheme that emerged during our interview analysis was the feeling of frustration, particularly with the setup of the device. Participants did not prefer products that left them feeling frustrated. An older adult expressed their feelings when comparing a few electronic medication adherence products in terms of the complexity of the setup process:

> *This one [product] was fine too. I found it a little frustrating, the mode and the hour getting it set in holding it. A little bit frustrating. The container holding the pills was there it was clear and easy to see. And I liked that. Setting it was a little bit of annoyance.* [Older adult 003]

One of the health care providers also shared similar thoughts:

> *It was just difficult [and] confusing to set. And also, it's a nightmare for the pharmacist. You'd have to have perfect intervals which is annoying. And that [electronic medication adherence product] then requires whoever's using it to take a dose to then set the right interval. What if, you know, they set 12-hour intervals, but they don't want to wake up at 6am. It could be very annoying.* [Pharmacist 023]

Overall, the study participants preferred electronic medication adherence products that provide them with a sense of assurance and privacy and those that do not make them frustrated.

Affordability

Another subtheme that was identified during the interview analysis was the affordability of a product, although this area was not something that was probed for as part of our semistructured interview guide. Many of the older adults mentioned being on a fixed income and, as such, expressed interest and concern regarding the cost of products. Health care providers expressed similar concerns, as shown in the following quote:

> *Cost may be a big issue because majority of my patients are under government coverage and asking them to pay a dollar is very, very difficult. And if you ask them to pay for something, unless it’s sponsored by the government or something, I don’t foresee us getting a patient to buy this. They’ll say they can take their medication in which we know they can’t. But still they will not pay to buy this.* [Pharmacist 020]

Physical and Cognitive Capability

Many participants mentioned that the physical and cognitive ability to use electronic medication adherence products was a crucial aspect when considering the use of a product. The use of electronic medication adherence products can be impacted by a user’s ability to access the medication from the device without any difficulties. Some products required manual dexterity to operate, including accessing medications from the device and setting the alarm. For example, individuals are required to press small buttons, flip switches, or rotate portions of the device to operate the product optimally. Similarly, good visual or hearing abilities are required to respond to audio-visual reminders. Some of these products also necessitate the significant cognitive capacity among users to understand how to operate and respond to alarms and prompts. These concerns were shared across all stakeholder groups. One older adult expressed the following concerns:

> *Some of the buttons were very hard to manipulate. They hurt your fingers or they're too small.* [Older adult 014]

Health care providers were also concerned about their patients experiencing difficulty using electronic medication adherence products, especially if they had an impairment or medical condition that could affect their manual dexterity. As one of the health care providers quoted:

> *The little one. It’s just so hard. It’s not hard for me to open it, but it’s small. It [would be] hard for an older adult that has some peripheral neuropathy or arthritis or something that would make it difficult or not useful to use this.* [Occupational therapist 038]

Technology Literacy and Learnability

Stakeholders felt that they needed to acquire new information related to technology to use some products adequately. This was problematic for some stakeholders because of the nonfamiliarity with technology and the technological complexity of the product. One health care provider expressed their concern through the following quote:

> *I do like the fact that there is a phone reminder, but then the person using it has to be tech-savvy as well. I don't know...I mean there are more older adults now who are able to use smart phones but the majority of my patients over the age of 80, they can barely use any smart device.* [Pharmacist 020]

Discussion

Principal Findings

This study highlights how the factors of an electronic medication adherence product along with the user factors impact the preference of a stakeholder to choose a product for medication management. For daily use of a product, users must be able to learn how to set up and regularly use the product appropriately. If the process of learning how to use the product is complex, it may impact the usability of the product, leading to improper product use. Usability can be defined as the manner in which a user interacts with a product and the ability of the user to perform the required tasks to use the product [18]. In many cases, electronic medication adherence products are advertised by manufacturers as simple to use and generally do not indicate whether the help of a health care provider or a caregiver is required. The usability study conducted as part of this larger mixed methods study showed that the usability and workload of electronic medication adherence products vary greatly among stakeholders and highlighted the need to assess these factors to provide guidance to health care providers regarding product recommendations to their patients [14].

Technology plays a significant role in every aspect of life [19,20]. Electronic medication adherence products are innovative products that may require some familiarity with technology. Research has shown that the interaction of older adults with technology is not similar to the interaction of younger adults and children with technology because of age-related cognitive and physical changes [21]. Individuals who are facing challenges in these areas, lack familiarity with technology, or feel uncomfortable relying on technology may dislike or be anxious about adopting a technology-based solution [21]. This uneasiness was expressed by older adults who participated in this study. However, with the continued use of a product, the technical aspects may become less challenging and this uneasiness may decrease.

In addition to product usability, this study showed that stakeholders prefer electronic medication adherence products that accommodate complex medication regimens, incorporate alarms, are secure and portable, require minimal technology use, and are affordable. Older adults with multiple comorbidities often take complex therapy regimens on a regular basis [8]. Managing multiple medications at variable times during the day is a complex task and often leads to confusion and improper medication use [8,12]. Many older adults in this study noted how some products would not accommodate their daily medication regimen and, therefore, indicated that the storage capacity of a device would be an important factor when choosing an electronic medication adherence product. In addition, patients may forget to take their medications or take them inappropriately because of age-related cognitive impairment [22,23]. To address this factor, numerous interventions have been developed, including alarms and audio-visual reminders [24]. These reminders include auditory sounds, vibrations, and/or flashing lights. Studies have reported that electronic medication packaging devices with audio-visual alarms or reminder systems can positively impact medication adherence [23,25]. According to this study, although an alarm or reminder function was considered to be important when choosing an electronic medication adherence product, the effort required to set up the alarm should also be considered. Therefore, to integrate electronic medication adherence products into daily use, electronic medication adherence products with reminder or alarm functions should be user-friendly, simple to set up, and easy to operate.

Studies have also shown that users prefer devices that are small in size and those that do not attract attention when used in public [24,25]. Previous studies have indicated that older adults prefer medication management products that are portable and do not interfere with their ability to leave the house [8]. Similar results were shown by our study participants, as they preferred electronic medication adherence products that could be easily carried or were less disruptive when leaving the home as they could bring their medications with them during day trips for outings with family and friends.

In general, older adults live on fixed or low incomes and, thus, may have difficulties affording an electronic medication adherence product. Some electronic medication adherence products have lower costs, whereas others range from a few hundred to thousand dollars. Although these products are being manufactured and marketed for improving medication adherence, the lack of reimbursement by provincial or private health care insurance plans prevents older adults from using these products. Studies have shown that people who are unable to afford costs for their medications are more likely to be nonadherent to long-term therapies for chronic medical conditions such as asthma, diabetes, heart failure, and depression [26,27]. Similarly, individuals may be reluctant to pay out of their pockets for a medication adherence aid that they may not be able to afford. Therefore, the cost of the electronic medication adherence product is another important feature to consider when choosing these products.

It is likely that there is no one product that will meet the needs of every older adult. Future iterations of and developments in electronic medication adherence products should consider the characteristics of patients when designing products. For example, the features of an electronic medication adherence product required to meet the needs of a 75-year-old patient with a history of arthritis, hypertension, and stroke who has limited vision and loss of function and is on a complex medication regimen with multiple drugs are considerably different from the needs of a highly functioning 65-year-old patient who is forgetful about once daily dosing of their antihypertensive medication. Caregivers may prefer a product that provides restricted access to medications so that their care recipient is not at risk of medication overdose, whereas another caregiver may want a device that automatically dispenses the medication and has a loud audio-visual alarm for their care recipient. Similarly, a 60-year-old person with Parkinson disease who is taking medications 6 times per day will find the storage capacity of an electronic medication adherence product to be a far more important factor when it comes to selecting an electronic medication adherence product. An individualized approach should be used to select a particular electronic medication adherence product, depending on the needs of the patient, to gain its full benefit for medication management. Manufacturers developing these products should collaborate with patient partners to address the needs of the population.

Strengths and Limitations

A major strength of this study was the involvement of multiple stakeholders in testing and providing feedback on electronic medication adherence products. Another strength was the inclusion of a wide range of electronic medication adherence products, from simple alarm-based pillboxes to highly sophisticated automated dispensers. As participants were interviewed after testing 5 products, rather than after each product, participants had the option of comparing all 5 electronic medication adherence products in their interviews. As a result, a limitation of this study was that participant responses may have been driven by their experience with 1 or 2 products, rather than discussing all products tested. To combat this, researchers ensured all products were visible during the interview to allow for better recall and prompted for additional details if no product was mentioned during the interview. Another limitation of this study was that interview data from stakeholder groups were not analyzed separately and, therefore, do not provide a detailed report of each group’s feedback for product preference.

Conclusions

In conclusion, this study demonstrated that preference for a particular electronic medication adherence product depends on multiple factors, including, but not limited to, the storage capacity, security, cost, and size of the device. Just one electronic medication adherence product will not be suitable for everyone. Therefore, health care providers should consider patient-related factors such as cognitive and functional capability to operate a device, medication regimens, and product features to choose the right product for the right patient. The manufacturers of these electronic medication adherence products should also consider the involvement of users in the beginning stages of product development for these technologies to ensure high acceptability, user friendliness, and affordability for end users. Policy makers should consider subsidizing the cost of electronic medication adherence products to make them affordable for people who are chronically ill and are on long-term therapies, allowing for a reduction in costs related to nonadherence in the health care system.

Acknowledgments

This study was supported by the Spark Program at the Centre for Aging and Brain Health Innovation. The authors would like to thank all the participants in the study for their time and input. The authors are grateful to Catherine Lee for assistance in copy editing and formatting the manuscript and Ruchi Parikh for a quantitative review of demographic data.

Conflicts of Interest

None declared.

References

1. United Nations Departments of Economic and Social Affairs. World Populations Prospect 2019: Highlights. New York, USA: United Nations; 2019.
2. Drug Use Among Seniors in Canada. Canadian Institute for Health Information. 2016. URL: https://www.cihi.ca/sites/default/files/document/drug-use-among-seniors-2016-en-web.pdf [accessed 2019-07-30]
3. Yee JL, Hasson NK, Schreiber DH. Drug-related emergency department visits in an elderly veteran population. Ann Pharmacother 2005 Dec;39(12):1990-1995. [doi: 10.1345/aph.1E541] [Medline: 16288080]
4. Pasina L, Brucato AL, Falcone C, Cucchi E, Bresciani A, Sottocorno M, et al. Medication non-adherence among elderly patients newly discharged and receiving polypharmacy. Drugs Aging 2014 Apr;31(4):283-289. [doi: 10.1007/s40266-014-0163-7] [Medline: 24604085]
5. Atkin PA, Finnegan TP, Ogle SJ, Shenfield GM. Functional ability of patients to manage medication packaging: a survey of geriatric inpatients. Age Ageing 1994 Mar;23(2):113-116. [doi: 10.1093/ageing/23.2.113] [Medline: 8023717]
6. Pantuzza L, Ceccato MG, Silveira M, Junqueira LM, Reis AM. Association between medication regimen complexity and pharmacotherapy adherence: a systematic review. Eur J Clin Pharmacol 2017 Nov;73(11):1475-1489. [doi: 10.1007/s00228-017-2315-2] [Medline: 28779460]
7. Advinha AM, Lopes MJ, de Oliveira-Martins S. Assessment of the elderly's functional ability to manage their medication: a systematic literature review. Int J Clin Pharm 2017 Feb;39(1):1-15. [doi: 10.1007/s11096-016-0409-z] [Medline: 27942949]
8. Reeder B, Demiris G, Marek KD. Older adults' satisfaction with a medication dispensing device in home care. Inform Health Soc Care 2013 Sep;38(3):211-222 [FREE Full text] [doi: 10.3109/17538157.2012.741084] [Medline: 23323721]
9. Anderson RE, Birge SJ. Cognitive dysfunction, medication management, and the risk of readmission in hospital inpatients. J Am Geriatr Soc 2016 Jul;64(7):1464-1468. [doi: 10.1111/jgs.14200] [Medline: 27305636]

10. Hoffmann C, Schweighardt A, Conn KM, Nelson D, Barbano R, Marshall F, et al. Enhanced adherence in patients using an automated home medication dispenser. J Healthc Qual 2018;40(4):194-200. [doi: 10.1097/JHQ.0000000000000097] [Medline: 28749791]
11. Patel T, Farooqi M, Carter C. Electronic medication adherence technologies: classification to guide use in older adults. Can Pharm J 2017 Jul 7;150(4):S26 [FREE Full text] [doi: 10.1177/1715163517719855]
12. Hall J, Bond C, Kinnear M, McKinstry B. Views of patients and professionals about electronic multicompartment medication devices: a qualitative study. BMJ Open 2016 Oct 17;6(10):e012915 [FREE Full text] [doi: 10.1136/bmjopen-2016-012915] [Medline: 27798025]
13. Look KA, Stone JA. Medication management activities performed by informal caregivers of older adults. Res Social Adm Pharm 2018 May;14(5):418-426 [FREE Full text] [doi: 10.1016/j.sapharm.2017.05.005] [Medline: 28528023]
14. Patel T, Ivo J, Faisal S, McDougall A, Carducci J, Pritchard S, et al. A prospective study of usability and workload of electronic medication adherence products by older adults, caregivers, and health care providers. J Med Internet Res 2020 Jun 2;22(6):e18073 [FREE Full text] [doi: 10.2196/18073] [Medline: 32348292]
15. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
16. Green J, Thorogood N. In: Seaman J, editor. Qualitative Methods for Health Research. Fourth Edition. London, UK: Sage Publication Ltd; 2018.
17. Sentiment. Merriam-Webster Dictionary. URL: https://www.merriam-webster.com/dictionary/sentiment [accessed 2019-12-01] [WebCite Cache ID https://www.merriam-webster.com/dictionary/sentiment]
18. Liu X, Blaschke T, Thomas B, De Geest S, Jiang S, Gao Y, et al. Usability of a medication event reminder monitor system (MERM) by providers and patients to improve adherence in the management of tuberculosis. Int J Environ Res Public Health 2017 Sep 25;14(10):- [FREE Full text] [doi: 10.3390/ijerph14101115] [Medline: 28946683]
19. Hedman A, Nygård L, Almkvist O, Kottorp A. Amount and type of everyday technology use over time in older adults with cognitive impairment. Scand J Occup Ther 2015 May;22(3):196-206. [doi: 10.3109/11038128.2014.982172] [Medline: 25580938]
20. Aldeer M, Javanmard M, Martin R. A review of medication adherence monitoring technologies. Appl Syst Innov 2018 May 6;1(2):14. [doi: 10.3390/asi1020014]
21. Charness N, Boot WR. Aging and information technology use. Curr Dir Psychol Sci 2009 Oct;18(5):253-258. [doi: 10.1111/j.1467-8721.2009.01647.x]
22. Hugtenburg J, Timmers L, Elders PJ, Vervloet M, van Dijk L. Definitions, variants, and causes of nonadherence with medication: a challenge for tailored interventions. Patient Prefer Adherence 2013;7:675-682 [FREE Full text] [doi: 10.2147/PPA.S29549] [Medline: 23874088]
23. Paterson M, Kinnear M, Bond C, McKinstry B. A systematic review of electronic multi-compartment medication devices with reminder systems for improving adherence to self-administered medications. Int J Pharm Pract 2017 Jun;25(3):185-194. [doi: 10.1111/ijpp.12242] [Medline: 26833669]
24. Vervloet M, Linn AJ, van Weert JC, de Bakker DH, Bouvy ML, van Dijk L. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. J Am Med Inform Assoc 2012;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
25. Checchi KD, Huybrechts KF, Avorn J, Kesselheim AS. Electronic medication packaging devices and medication adherence: a systematic review. J Am Med Assoc 2014 Sep 24;312(12):1237-1247 [FREE Full text] [doi: 10.1001/jama.2014.10059] [Medline: 25247520]
26. Piette JD, Heisler M, Wagner TH. Problems paying out-of-pocket medication costs among older adults with diabetes. Diabetes Care 2004 Feb;27(2):384-391. [doi: 10.2337/diacare.27.2.384] [Medline: 14747218]
27. Piette JD, Heisler M, Wagner TH. Cost-related medication underuse among chronically ill adults: the treatments people forgo, how often, and who is at risk. Am J Public Health 2004 Oct;94(10):1782-1787. [doi: 10.2105/ajph.94.10.1782] [Medline: 15451750]

Edited by G Eysenbach; submitted 01.02.20; peer-reviewed by J Chung, B Pouls, J Manning, A Carcone; comments to author 29.06.20; revised version received 17.08.20; accepted 11.11.20; published 01.12.20.

Please cite as:
Faisal S, Ivo J, McDougall A, Patel T
Stakeholder Feedback of Electronic Medication Adherence Products: Qualitative Analysis
J Med Internet Res 2020;22(12):e18074
URL: https://www.jmir.org/2020/12/e18074
doi:10.2196/18074
PMID:33258778

Original Paper

Screening for Depression in Daily Life: Development and External Validation of a Prediction Model Based on Actigraphy and Experience Sampling Method

Olga Minaeva[1], MSc; Harriëtte Riese[1], PhD; Femke Lamers[2], PhD; Niki Antypa[3], PhD; Marieke Wichers[1], PhD; Sanne H Booij[4,5], PhD

[1]Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

[2]Department of Psychiatry, Amsterdam UMC, Amsterdam Public Health Research Institute, Vrije Universiteit, Amsterdam, Netherlands

[3]Department of Clinical Psychology, Institute of Psychology, Leiden University, Leiden, Netherlands

[4]Interdisciplinary Center for Psychopathology and Emotion regulation, Department of Developmental Psychology, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, Netherlands

[5]Center for Integrative Psychiatry, Lentis, Groningen, Netherlands

Corresponding Author:
Olga Minaeva, MSc
Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), Department of Psychiatry
University Medical Center Groningen
University of Groningen
Hanzeplein 1
Groningen, 9713 GZ
Netherlands
Phone: 31 50 361 2065
Email: o.minaeva@umcg.nl

Abstract

Background: In many countries, depressed individuals often first visit primary care settings for consultation, but a considerable number of clinically depressed patients remain unidentified. Introducing additional screening tools may facilitate the diagnostic process.

Objective: This study aimed to examine whether experience sampling method (ESM)-based measures of depressive affect and behaviors can discriminate depressed from nondepressed individuals. In addition, the added value of actigraphy-based measures was examined.

Methods: We used data from 2 samples to develop and validate prediction models. The development data set included 14 days of ESM and continuous actigraphy of currently depressed (n=43) and nondepressed individuals (n=82). The validation data set included 30 days of ESM and continuous actigraphy of currently depressed (n=27) and nondepressed individuals (n=27). Backward stepwise logistic regression analysis was applied to build the prediction models. Performance of the models was assessed with goodness-of-fit indices, calibration curves, and discriminative ability (area under the receiver operating characteristic curve [AUC]).

Results: In the development data set, the discriminative ability was good for the actigraphy model (AUC=0.790) and excellent for both the ESM (AUC=0.991) and the combined-domains model (AUC=0.993). In the validation data set, the discriminative ability was reasonable for the actigraphy model (AUC=0.648) and excellent for both the ESM (AUC=0.891) and the combined-domains model (AUC=0.892).

Conclusions: ESM is a good diagnostic predictor and is easy to calculate, and it therefore holds promise for implementation in clinical practice. Actigraphy shows no added value to ESM as a diagnostic predictor but might still be useful when ESM use is restricted.

(J Med Internet Res 2020;22(12):e22634) doi:10.2196/22634

KEYWORDS

actigraphy; activity tracker; depression; experience sampling method; prediction model; screening

Introduction

Depressive disorders represent a major public health concern as they are the most prevalent psychiatric disorders and a leading cause of disability worldwide [1,2]. In many countries, depressed individuals often first visit primary care settings when they seek help [3]. Even though most nondepressed individuals can be accurately excluded in primary care [4], a considerable number of clinically depressed patients remains unidentified [5]. Thus, general practitioners can correctly identify between 41.7% and 53.0% of cases of depression with a sensitivity between 41.3% and 59.0% and a specificity between 74.5% and 87.3% [4]. An additional challenge in the detection of depression arises because patients often present with undefined or somatic illness [6], resulting in depression going undetected and often untreated for a longer time period [7].

According to a meta-analysis on the clinical diagnosis of depression in primary care, the accuracy of identification of depression can be improved by prospective examination over an extended period [4]. Therefore, introducing additional screening tools that allow continuous monitoring during daily life may facilitate the diagnostic process and improve referral of depressed individuals to the right care providers. A good candidate for a screening tool that holds particular value for studying mood disorders is the experience sampling method (ESM) [8]. Most commonly delivered via a smartphone, ESM involves repeated, intensive sampling of respondents' current affect, experiences, and behaviors while they are engaged in their daily activities, in their natural environments [9]. Hypothetically, this might be an optimal way of detecting depression risk, as a person can repeatedly assess his/her affect and behaviors in daily life with minimal retrospective recall bias [8]. Previous studies have shown that higher levels of negative affect, which is commonly assessed with ESM, are strongly associated with depression [10]. Furthermore, this method comes closest to the advised method to do longitudinal assessments of depressive symptoms [11,12]. A problem with ESM, however, may be that it places too high a burden on the patient, leading to reduced compliance [9]. Hence, ESM as a screening tool may not be suitable for everyone, warranting the exploration of more passive ways of collecting data as well.

A potential candidate for depression screening that involves passive data collection is ambulatory assessment of actigraphy data from sensors, such as activity trackers. Such activity trackers are now widely used, and they provide ecologically valid data about behavior [13-16]. The data derived from activity trackers include patterns of sleep, physical activity, and circadian rest-activity rhythm (RAR). Alterations in these patterns have been found in depression [17-22] and contributed to objective differentiation of depression subtypes [23]. Further, these behavioral parameters are easily and passively measurable by actigraphy and do not require any invasive procedure or active participation from individuals. Therefore, they do not create an additional burden for an individual [15]. However, it is important to note that the alterations in these actigraphy patterns are not specific for depression, since altered sleep, physical activity, and RAR are present in many other health conditions [24,25]. Therefore, most likely, actigraphy may only be used in addition to other measures when screening for depression. However, the predictive value of actigraphy-based measures, alone and in combination with other measures, for depression remains to be examined.

Sleep, physical activity, and RAR domains are associated with each other [26,27]. However, previous researchers who found associations between actigraphy data and depression included measures from only 1 or 2 of these domains (ie, studying physical activity only, sleep only, or circadian RAR and sleep) [20,21,28-30], with a rare exception [31]. Therefore, probably not all actigraphy-based measures that have been previously associated with depression will have a unique predictive value as part of a multidimensional screening tool. Thus, it is currently still unclear which combination of actigraphy measures are most strongly associated with depression risk.

Using both ESM and actigraphy approaches together for continuous and everyday monitoring of behavioral and affective aspects in depression could be a promising screening tool. While it is assumed that self-reports of depression-related affect and behaviors as assessed with ESM predict depression better than behavioral sensor data, this assumption has never been tested before. To our knowledge, there is only 1 recent study that attempted to predict depression by using both ESM and actigraphy data; however, it was focused on the elderly, had a smaller sample size (N=47), and had no external validation [32]. Currently, it is not yet clear how these approaches perform and if they can be used for screening purposes, both separately and in combination.

In this study, we examined (1) whether ESM-assessed depression-related affect and behavior could discriminate between depressed and nondepressed individuals, (2) whether actigraphy data could discriminate between depressed and nondepressed individuals, and (3) whether actigraphy has added value with respect to the use of ESM. Therefore, we compared the performance of the prediction models with ESM only and actigraphy only to assess the added benefit of the individual domains and then evaluated the performance of the prediction model with both domains included. First, we hypothesized that ESM measures would have a better discriminating ability in distinguishing individuals with and without a diagnosis of depression than actigraphy measures. Second, adding actigraphy measures to the ESM prediction model would improve the discriminating ability. To test these hypotheses, we used 2 data sets for development and validation of the prediction models.

Methods

Study Population

We used data from the Netherlands Study of Depression and Anxiety (NESDA) [33] to develop prediction models, and data from the Mood and Movement in Daily Life (MOOVD) study to validate them [34].

Development Data Set

In short, NESDA is an ongoing multisite longitudinal cohort study among 2981 adults (aged 18 to 65 years) at baseline, including individuals with depressive and/or anxiety disorders and healthy control subjects, which were recruited from the general population. Details about the total NESDA sample are provided elsewhere [33]. In this study, we used a subsample from the Ecological Momentary Assessment and Actigraphy (EMAA) substudy, which combined 14 days of ESM (5 times a day) with continuous actigraphy [31,35]. A flowchart of the inclusion process for this study is provided in Multimedia Appendix 1. Individuals with a diagnosed episode of major depressive disorder and/or dysthymia in the past month (n=43) and individuals with no lifetime depressive or anxiety disorder (control group, n=82) based on the Composite International Diagnostic Interview (CIDI) [36] were included in the study. Severity of depressive symptoms was assessed with the self-reported Inventory of Depressive Symptomatology (IDS-SR) [37]; the mean IDS-SR score represents moderate depressive symptoms.

Validation Data Set

The MOOVD study is an ambulatory assessment study among matched depressed and nondepressed individuals (n=54; aged 20 to 50 years) [34]. Depressed individuals were recruited from 3 psychiatric outpatient centers; nondepressed individuals were recruited from the general population in the Netherlands. This study combined 30 days of ESM (3 times a day) with continuous actigraphy. Depressed individuals (n=27) with a major depressive episode at the time of the interview or within two months prior to the interview, according to the CIDI, were included. Nondepressed individuals (n=27) were free of any mood disorders at the moment of inclusion but were allowed to have a history of depression (n=1, >7 years ago). Severity of depressive symptoms was assessed with the Beck Depression Inventory-II (BDI-II) [38]; the mean BDI-II score represents severe depressive symptoms. Individual scores, however, range from no/mild to severe depressive symptoms in both data sets (IDS-SR score between 9 and 64; BDI-II score between 15 and 51).

Actigraphy Assessments

NESDA participants wore the wrist-worn GENEActiv accelerometer (Activinsights Ltd) for 24 hours a day for 14 days. GENEActiv validity studies have demonstrated strong correlations for criterion validity (Pearson *r*=0.79-0.98) [39] and a good ability to determine sedentary behavior in adults (aged 18 to 55 years) (Pearson *r*=0.81) [40]. Details of the actigraphy measurements of the NESDA-EMAA substudy are provided elsewhere [31]. MOOVD participants were assessed with an Actical accelerometer (Respironics, Inc) for 24 hours a day for 30 days. In the laboratory study, the Actical demonstrated high reliability (intraclass correlation coefficient=0.92) and validity (*r*=0.81) in adolescents [41]. More information about the actigraphy assessments of the MOOVD study can be found elsewhere [42].

Experience Sampling Methodology (ESM)

In NESDA, participants took part in the ESM assessment for 2 weeks, during which they filled out questions on smartphones 5 times a day. The electronic diary had a fixed design with 3 hours between each beep, and the questionnaire included items on current mood states, social interactions, daily experiences, and behaviors [35]. Of all ESM assessments of all participants, only 8.3% were missing and all included participants had enough valid data points (>60 time points). In the MOOVD study, participants completed questionnaires on an electronic diary, the PsyMate (PsyMate BV) [43]. The electronic diary had a fixed design with 3 beeps a day, 6 hours apart. The electronic questionnaire contained items about mood, sleep, activities, as well as social interactions, important events, rumination, and self-esteem. Detailed information about the ambulatory assessment procedure is provided elsewhere [34]. All included participants had enough valid diary measurements (>60 time points).

Outcome Variables

As the main outcome measure for both data sets, we used presence or absence of a diagnosis of depression (major depressive disorder and/or dysthymia) based on DSM-IV criteria [44], assessed with the Composite International Diagnostic Interview (CIDI), version 2.1. The CIDI is a fully structured interview designed for assessing mental disorders according to the diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) [36]. All CIDIs were performed by well-trained research assistants, mainly psychologists, mental health care nurses, and residents in psychiatry. In the development data set, participants were diagnosed with the CIDI instrument during the regular NESDA interview wave, which was a maximum of 31 days prior to the actigraphy and ESM assessments. In the validation data set, participants started the actigraphy and ESM assessments immediately following the screening CIDI.

Predictor Variables

Objectively assessed sleep, physical activity, and computed RAR variables, as calculated from the actigraphy data, and ESM-assessed depression-related affect and behavior were used as predictors in our models. Preprocessing of the raw actigraphy data was done in R using GGIR package version 1.5-18 (https://cran.r-project.org/web/packages/GGIR/) for the NESDA data set and described elsewhere [31]. Almost all variables were created similarly in the NESDA and the MOOVD data sets; any exceptions are mentioned.

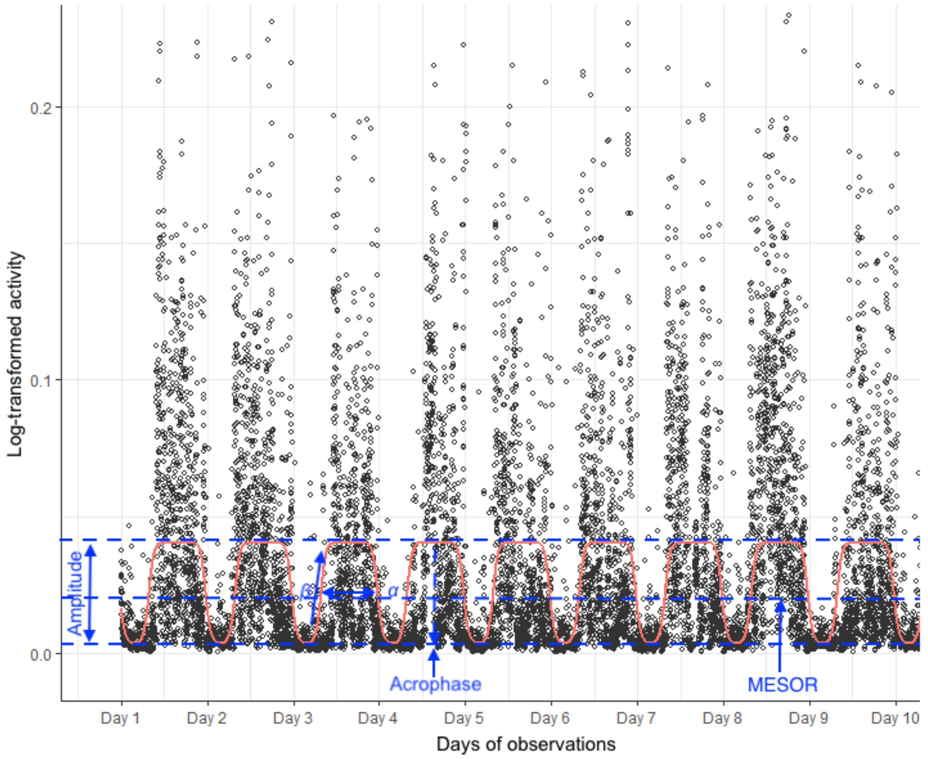
In the NESDA data set, physical activity was assessed as gross motor activity per day and as minutes of moderate-to-vigorous physical activity (MVPA) per day [45]. Objective gross motor activity was estimated by calculating the Euclidian Norm Minus One (ENMO) per individual per day [31]. Based on those calculations, average estimates of gross motor activity were estimated for each participant. To keep consistent with earlier papers, MVPA was defined as ENMO values greater than 125 mg [31]. In the MOOVD data set, the Actical actigraphy device did not allow the extraction of the raw actigraphy data (data in SI units represented as acceleration in x, y, and z axes), and therefore the data were not processed with the GGIR package. Instead, activity counts (AC) and activity intensity with 4 categories (sedentary, light, moderate, and vigorous) were calculated as a measure of motor activity by an in-built algorithm of the Actical software.

Sleep was assessed as total sleep time (TST) in hours and sleep efficiency per night (%) [21,46]. In the NESDA data set, TST was estimated with the GGIR package and equaled the accumulated nocturnal sustained inactivity bouts. Sleep efficiency was calculated as TST divided by time in bed (estimated by the GGIR package). For the MOOVD data set, TST was calculated as the sum of estimated sleep periods based on the Sadeh algorithm [47]. Sleep efficiency was calculated as a percentage of time scored as sleep during the time spent in bed.

To characterize circadian RAR profiles, individual actigraphy data sets were fitted to an extended cosinor model [48] using nonlinear least-squares regression (RAR package version 1.0.0 for R). This allowed the estimation of 5 circadian curve parameters for each participant, namely the midline estimating statistic of rhythm (MESOR), amplitude, acrophase, α, and β, as well as the circadian rhythmicity index (*F* statistic) (see Figure 1 for more details).

To assess depression-related affect and behavior, we selected ESM items that, in terms of content, matched DSM-V diagnostic criteria for depression. For example, the symptom of “sad or depressed mood” could be represented by the momentary affect state, “I feel sad.” The following items that were present in both data sets were included: sad or depressed mood, irritation, appetite, energy, tiredness, loss of interest, enthusiasm, guilt, concentration, and sleep disturbances. A complete list of included ESM items from the NESDA and the MOOVD data sets can be found in Multimedia Appendix 2. All items were scored on 7-point Likert scales. The sum score of these items calculated for each day and then averaged across 14 days represented depression-related affect and behavior. The person-level Cronbach α for depression-related affect and behavior was .928 in the NESDA data set and .936 in the MOOVD data set. Gender, age, and education level were included in the analysis as covariates.

Figure 1. Example of rest-activity rhythm parameters derived from the extended cosinor model. The midline estimating statistic of rhythm (MESOR) is a mean of the modeled activity curve; amplitude is the difference between the peak and trough of the fitted curve, herein estimating the range of activity levels across the 24-hour period; acrophase is a phase marker indicating the time when the fitted curve reaches its peak (ie, time of maximal activity levels across the 24-hour period); α is the relative width of the curve at the middle of the peak; β is an indicator of the steepness of the rise and fall of the curve; and circadian rhythmicity index (F statistic) is an indicator of the strength of circadian rhythmicity (a goodness-of-fit measure for which higher values indicate smaller discrepancies between actigraphy data and values predicted by the cosinor model).



Statistical Analysis

Multicollinearity for all predictor variables was checked by calculating Spearman correlations and the variance inflation factors (VIFs). Spearman correlations above 0.80 and VIFs above 10 were considered to be indicative of severe collinearity [49,50]. In this situation, 1 of the collinear variables that was the least related to the outcome variable was removed from further analysis. Fractional polynomials were used to check the presence of nonlinear associations of the continuous predictors to the outcome variable. Cubic association was found for sleep duration (TST) and therefore was included as such in the analysis.

Building Single-Domain Models

The next step was to build single-domain models for actigraphy and ESM measurements separately. For the ESM model **(1)**, we included the ESM "depression-related affect and behavior" score and covariates (age, gender, and education), as their association to depression has been consistently shown [51,52] and this information can be easily added to a screening tool. The sum score was chosen instead of including the ESM items in the analysis separately, as it was meant to mimic depressive symptoms.

Group status = a_0 + $a_1$sum score + $a_2$age + $a_3$gender + $a_4$education **(1)**

where "a" represents the regression coefficients from the model: a_1-a_4 are predictor coefficients and a_0 is the intercept.

To build an actigraphy model, we used a multivariate backward stepwise logistic regression approach [53]. A baseline actigraphy model included all actigraphy predictors (amplitude, acrophase, α, β, *F* statistic, ENMO, TST, and sleep efficiency) as well as predefined covariates (age, gender, and education). The MESOR was found to be collinear with ENMO and the least related to the outcome variable; therefore, it was removed from the further analysis. Since different physical activity metrics (ENMO and AC) were available for the development and the validation data sets, we standardized ENMO and AC to alleviate a comparison of the actigraphy models in two data sets. In the following steps, we removed the least significant actigraphy variable (with the highest *P* value) and compared the Akaike information criterion (AIC) value to the AIC from the previous model. A significantly smaller AIC indicates a better model. The procedure was repeated until we defined the optimal combination of actigraphy predictors based on the AIC. The regression equation for the final actigraphy model is included below **(2)**:

Group status = a_0 + $a_1$standardized ENMO + $a_2$acrophase + $a_3$age + $a_4$gender + $a_5$education **(2)**

Building a Combined-Domains Model

A multivariable backward stepwise logistic regression model was performed to examine what combination of predictors (ie, actigraphy, ESM) resulted in the optimal prediction model for distinguishing between depressed and nondepressed individuals. Since we used a backward approach, the baseline model included all predictors with unique information based on the single-domain models (the actigraphy and the ESM models). In the following models, we used a procedure where we removed variables one by one, based on the highest *P* value, and checked every time whether AIC improved until we defined the final prediction model based on the AIC [54]. The regression equation for the final combined-domains model is included below **(3)**:

Group status = a_0 + $a_1$sum score + $a_2$standardized ENMO + $a_3$age + $a_4$gender + $a_5$education **(3)**

Evaluation and Validation of the Single-Domain and the Combined-Domains Prediction Models

To evaluate the performance of the combined-domains model and the single-domain models, we utilized goodness-of-fit indices and calibration curves and assessed the discriminative ability of the models (the area under the receiver operating characteristic curve [AUC]) [55]. The goodness-of-fit indices and calibration curves evaluate how close the predicted and observed estimates are. The AUC represents the ability of the models to distinguish between patients with and without the depression diagnosis and ranges from 0.5 (by chance) to 1.0 (perfect discrimination). These quality indicators of all 3 models were compared with a basic model with only the covariates and to each other. As suggested in the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement [56], we performed bootstrapping techniques for internal validation of the model to simulate the performance of the prediction model in comparable patient data sets.

As a next step, we performed an external validation by using the developed single-domain and combined-domains models from the NESDA data set to assess the predictive performance of the models in the validation sample (ie, the MOOVD data set), calculating the discriminative ability reflected by the AUC. The NESDA data set was chosen to be a development data set because it was larger than the MOOVD data set, to minimize the possibility of overfitting while building the prediction model. The MOOVD data set was suitable for the external validation because inclusion criteria and measurements of depression were similar. Again, we compared the final combined-domains prediction model with the actigraphy and ESM models in the MOOVD data set to check whether it still had a better fit than the single-domain models. The results were reported according to the TRIPOD statement [56].

Results

The characteristics of the development (NESDA) and validation (MOOVD) data sets are given in Table 1.

Table 1. Characteristics of development and validation data sets.

	NESDA[a] (n=125)		MOOVD[b] (n=54)	
	Depressed (n=43)	Control (n=82)	Depressed (n=27)	Control (n=27)
Study characteristics				
Data collection period	2014-2017		2012-2014	
Setting	General population, primary health care, and mental health care		General population, and outpatient centers for psychiatry	
Inclusion criteria for cases	Receiving a depression diagnosis 1 month before the ESM[c]/actigraphy assessment		Receiving a depression diagnosis 1 month before the ESM/actigraphy assessment	
Outcome	Presence or absence of a depression diagnosis (MDD[d] and/or dysthymia) based on DSM-IV[e] criteria		Presence or absence of a depression diagnosis (MDD and/or dysthymia) based on DSM-IV criteria	
Prevalence of outcome, n (%)	43 (34.4)		27 (50.0)	
Sociodemographic characteristics				
Female, n (%)	29 (67.4)	46 (56.1)	20 (74.1)	20 (74.1)
Age, mean (SD)	52.14 (9.57)	51.50 (12.70)	34.70 (9.86)	34.04 (8.96)
Education (high), n (%)	13 (30.2)	49 (59.8)	17 (63.0)	19 (70.4)
Psychopathology				
Depression severity instrument	IDS-SR[f]		BDI-II[g]	
Depression severity, mean (range)	34.53 (9-64)	5.44 (0-25)	31.33 (15-51)	2.26 (0-10)
AD[h] and/or BD[i] use, n (%)	23 (53.5)	4 (4.9)	15 (55.6)	1 (3.7)

[a] NESDA: the Netherlands Study of Depression and Anxiety.

[b] MOOVD: Mood and Movement in Daily Life.

[c] ESM: experience sampling method.

[d] MDD: major depressive disorder.

[e] DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, fourth edition.

[f] IDS-SR: Inventory of Depressive Symptomatology (self-report).

[g] BDI-II: Beck Depression Inventory-II.

[h] AD: antidepressant.

[i] BD: benzodiazepine.

Predictors of depression in the final models were the “depression-related affect and behavior” sum score for the ESM model and gross motor activity (ENMO) and the time of maximal activity levels across the 24-hour period (acrophase) for the actigraphy model (Table 2). The combined-domains prediction model included the “depression-related affect and behavior” sum score and ENMO variables.

For the ESM model, the predictive capacity was 95.2%, which is 29.6% higher than in the null (only intercept) model (65.6%). Calibration of the ESM model was adequate with a Nagelkerke R^2 statistic of 0.904 and a Hosmer-Lemeshow goodness-of-fit test of 0.960 (P=.998).

For the actigraphy model, the predictive capacity of the final step of the backward selection model was 71.8% (6.2% higher than the null model). Calibration of the actigraphy model was adequate with a Nagelkerke R^2 statistic of 0.357 and a Hosmer-Lemeshow goodness-of-fit test of 10.678 (P=.221).

The final (combined-domains) model had the same predictive capacity as the ESM model (95.2%). Calibration of the final model was adequate with a Nagelkerke R^2 statistic of 0.913 and a Hosmer-Lemeshow goodness-of-fit test of 1.786 (P=.987). In all 3 calibration plots, the slope approached the diagonal (see Multimedia Appendix 3).

Table 2. Predictors of depression included in the experience sampling method (ESM) prediction model, the actigraphy prediction model, and the final combined-domains model.

Predictor	B[a]	SE	*P* value	Exp(B)[b]	95% CI for Exp(B)
ESM prediction model					
Intercept	–27.880	8.668	.001[c]	0.000	
Gender	–2.955	1.597	.06	0.052	0.002-1.191
Age	0.143	0.068	.03[c]	1.154	1.009-1.319
Education	–1.701	1.058	.11	0.182	0.023-1.452
Sum score[d]	0.891	0.243	<.001[c]	2.437	1.513-3.923
Actigraphy prediction model					
Intercept	0.050	2.954	.99	1.052	
Gender	–1.021	0.513	.046[c]	0.360	0.132-0.984
Age	–0.012	0.020	.54	0.988	0.950-1.027
Education	–1.711	0.456	.001[c]	0.181	0.074-0.441
Z-score ENMO[e]	–1.004	0.283	<.001[c]	0.367	0.211-0.638
Acrophase	0.289	0.177	.10	1.335	0.943-1.888
Final combined-domains prediction model					
Intercept	–25.841	9.314	.01[c]	0.000	
Gender	–3.751	1.740	.03[c]	0.024	0.001-0.712
Age	0.102	0.076	.18	1.107	0.953-1.286
Education	–1.988	1.163	.09	0.137	0.014-1.339
Sum score	0.919	0.263	<.001[c]	2.508	1.498-4.197
Z-score ENMO	–1.117	0.767	.15	0.327	0.073-1.472

[a] B: regression coefficient.

[b] Exp(B): exponentiation of the B coefficient (odds ratio).

[c] Significant *P* value (*P*<.05).

[d] Sum score: a person-level sum score of the ESM items that represent depression-related affect and behavior.

[e] ENMO: Euclidian Norm Minus One.

In the development data set, the discriminative ability in predicting depression was good in the actigraphy model (AUC=0.790) and excellent in the ESM model (AUC=0.991) and the combined-domains model (AUC=0.993) (Table 3, Figure 2). The estimations did not meaningfully change after internal validation. The discriminative ability of the model in the external validation sample was reasonable in the actigraphy model (AUC=0.648) and very good in the ESM model (AUC=0.891) and the combined-domains model (AUC=0.892) (Multimedia Appendix 4). Calibration of all models was adequate, with the slope approaching the diagonal in all 3 calibration plots (Multimedia Appendix 5). Multimedia Appendices 6 and 7 provide an overview of different probability thresholds and their respective classification measures (sensitivity, specificity, and predictive values) for all 3 models of interest in the development and validation data sets.

Table 3. The discriminative ability of the final combined-domains model and the basic, actigraphy, and experience sampling method (ESM) models.

Models	Development data set (NESDA[a]; n=125)		Validation data set (MOOVD[b]; n=54)	
	AUC[c]	95% CI	AUC	95% CI
Basic model[d]	0.704	0.610-0.798	0.464	0.304-0.624
ESM model	0.991	0.981-1.000	0.891	0.802-0.979
Actigraphy model	0.790	0.713-0.867	0.648	0.492-0.803
Combined-domains model	0.993	0.983-1.000	0.892	0.800-0.984

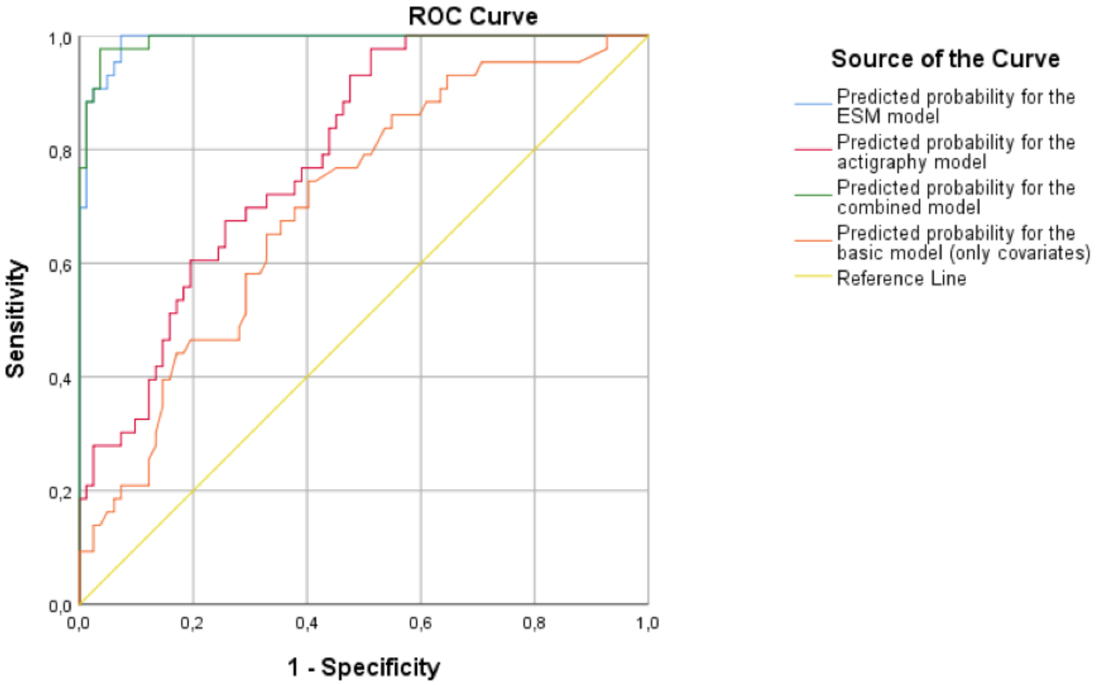
[a] NESDA: the Netherlands Study of Depression and Anxiety.

[b] MOOVD: Mood and Movement in Daily Life.

[c] AUC: area under the receiver operating characteristic curve.

[d] Basic model has covariates included only (gender, age, and education).

Figure 2. The receiver operating characteristic (ROC) curves of the basic model, the experience sampling method (ESM) model, the actigraphy model, and the combined-domains model in the development data set (the Netherlands Study of Depression and Anxiety).



Discussion

Principal Results

In this paper, we have developed 3 prediction models based on ESM measures of depression-related affect and behaviors, on actigraphy, and on their combination, for discriminating currently depressed from nondepressed individuals. To our knowledge, this is the first study that has created and compared such models for their individual performance and for their performance in combination, and using both internal and external validation to test their discriminative abilities. The ESM model had an excellent predictive potential in discriminating depressed and nondepressed individuals in both the development and the validation data sets. The actigraphy model, in turn, had a reasonable predictive potential alone but could not compete with the ESM model in predictive performance. The combined-domains prediction model, which included the ESM measure as well as the best combination of actigraphy measures, was very similar to the ESM model in performance in both the development and the validation data sets. Hence, from the results we can conclude that the ESM and actigraphy measures both have the potential to serve as an additional screening tool; however, actigraphy does not have added value when combined with ESM.

Comparison with Prior Work

The ESM sum score combined items about positive affect, negative affect, sleep, and appetite. Measuring these symptom-related items over a prolonged period of time resulted in successful discrimination between depressed and nondepressed individuals. The constructed prediction model performed excellently, not only in the development but also in the validation data set. In line with our findings, multiple previous studies showed correlations between related measures, mainly negative affect and positive affect (assessed with ESM), and depression [10,12,57]. Another study attempted to estimate depressive symptoms based on the ESM items and found significant correlations with depressive symptoms assessed with symptom questionnaires [11]. Despite the strong correlation between ESM-assessed depression-related affect and behavior and depression, there have been no diagnostic prediction studies using such ESM data. Existing literature, however, indicates that both negative affect and positive affect play a role in predicting relapse of depression [58,59] and future treatment outcome [60,61]. Additionally, negative affect was found to be predictive of depression onset in youth [62]. There are a very limited number of prognostic studies available, focusing on older adults [63]. Authors have identified the items “sad” and “tired” as sensitive measures that have the potential to predict future depression status in older adults. Although replication in other types of samples is warranted, our results suggest that the ESM measure has potential for screening purposes in clinical practice.

Interestingly, self-report measures of depressive affect and behavior largely overperformed objective measures of behavior in distinguishing depressed from nondepressed individuals. Researchers often make the implicit assumption that subjective measures are inaccurate and cannot compete with objective assessments. Self-reported affect and behavior, indeed, might not align with actual observed affect or behavior due to a difference between perceived and actual behavior [64]. Nevertheless, this bias in how a depressed person perceives their own emotions and feelings might be highly useful for predicting depression.

Although the actigraphy model showed good performance in the development data set, its performance dropped in the validation data set. While this discrepancy may be a signal of worse performance in other samples, it could also partly be due to differences in physical activity metrics in use. We found that metrics from the GENEActiv and Actical accelerometers (ENMO and AC, respectively) have never been directly compared before, hence we could not apply any known formula to transform data from one into the other. To overcome this issue, we z-transformed ENMO and AC before the analysis to adjust the variables to the same scale from –1 to 1. This step improved the performance of the actigraphy model, although it was still lower in the validation data set. Of note, it has previously been shown that the placement of the accelerometer device on the body influences its outcome measures in both the level of and fluctuations in activities [65]. Wrist placement was recommended as a basic technique to capture motor activity in depressed patients because it records whole-body movement and gestures [66]. In this study, the devices were worn on the wrist, which implies that the model is only applicable to actigraphy as measured using the wrist placement. Despite the existing challenges and a lower predictive capacity compared with ESM, actigraphy, being a passive data collection method, might still be useful for screening purposes in some cases, for example, in a situation where the use of ESM is not possible or too bothersome for particular individuals.

In agreement with previous studies that used actigraphy data [20-22,31,67], we have found strong associations between lower levels of physical activity and depressive symptoms/depressive disorders. Other RAR characteristics such as MESOR, amplitude, α, β, and F statistic, calculated with the extended cosinor analysis, were not sufficiently predictive when assessed together with the overall activity level (ENMO) and therefore, were eliminated from the model. Interestingly, the individual associations between these parameters and depression were significant; however, when these variables were combined, only a few remained in the model (ENMO and acrophase). The fact that only a few variables remained might be due to the potential overlap of these parameters. Even though the association between circadian rest-activity parameters and depression has been shown previously [30,68-71], to our knowledge, there are no studies that specifically compared the predictive ability of various actigraphy-based parameters from different domains in distinguishing depressed and non-depressed individuals. Hence, future studies should further examine to what extent these different variables are measuring partly the same concepts.

The combined-domains model had an excellent performance that was highly similar to the performance of the ESM model. Adding the actigraphy model component (ENMO or AC) did not significantly improve the performance of the combined-domains model. The explanation for this lack of improvement might be that ESM captures part of actigraphy variance. This overlap may be in part because both ESM and actigraphy assess sleep; one assesses subjective sleep duration and subjective sleep quality, while the other assesses the same characteristics objectively. Another example of shared variance might be that complaints like anhedonia could result in a patient being less active and hence, the reduced physical activity may be a result of the depressive symptoms. As in the previous example, concentration problems could be associated with lower sleep quality [72]. This makes questionable the hypothesis of whether actigraphy has added value in its own right when combined with ESM. To our knowledge, there is only 1 recent study that attempted to predict depression by using both ESM and actigraphy data; however, it was focused on community-dwelling older adults, had a smaller sample size (N=47), and had no external validation [32]. These researchers used a wide range of various predictors based on actigraphy and ESM in a machine learning approach to define the optimal combination of the predictors and build a prediction model. As in our study, these researchers developed their model on the basis of daily mean ESM scores, and actigraphy-based daily mean activity levels and daily sleep efficiency variables, although the latter was removed from our prediction models. Hence, the chosen variables showed predictive potential in detecting depression even when different selection approaches were applied. More details of the studies discussed in this paragraph are provided in Multimedia Appendix 8.

Strengths and Limitations

The main strength of this study was the external validation of the prediction models in a different and adequate data set that prevented overstating the results [56,73]. The fact that there were substantial differences between the 2 data sets and the results were consistent in both data sets provides strong evidence that the prediction models can be generalized to new patients [73]. Finally, all constructed models include variables that are ecologically valid and easy to measure in real life. There are also some limitations of the study that need to be mentioned.

First, different physical activity metrics (ENMO versus AC) limited the ability to externally validate the actigraphy prediction model in an optimal way. To avoid this problem in the future, researchers should preferably use the accelerometers that allow access to the raw data. In this case, the same output metric can be chosen so that the algorithms for the computation will be the same or at least comparable. Second, the development data set had a small number of individuals in the depression groups who reported no or mild depressive symptoms. This was most likely due to logistic reasons typical of large cohort studies, as our study was. Participants were diagnosed with the CIDI instrument during the regular NESDA interview wave, which was a maximum of 31 days prior to the actigraphy assessments, whereas depression severity assessment with the IDS-SR questionnaire was not necessarily done close to the actigraphy assessment period (up to 72 days prior and two cases of 74 and 351 days after the CIDI). This situation, however, could possibly deflate the association rather than inflate it, and the model would have more predictive capacity in more acutely depressed individuals. Third, the size of the validation data set was smaller than suggested by some simulation studies, which required at least 100 events for the validation sample. This suggestion, however, was based on limited simulation studies and it lacks the empirical evidence to guide research [56]. The sample size, therefore, is often determined by the available data, as was the case in our study [56]. Finally, both development and validation samples were collected in the Netherlands, which limits generalizability to countries with similar ethnicity profiles and health care systems.

Concerning future research, the constructed ESM models performed excellently with 14 and 30 days of continuous measurements in the development and validation data sets, respectively. However, if a shorter duration of ESM measurements showed similar performance in the prediction of depression, a shorter measurement period could potentially reduce the burden on patients who struggle with longer ESM regimes. Indeed, it has previously been shown that an association between negative affect and depression can already be detected using ESM for 6 to 7 days [60,74,75] or even 3 days [76]. Further, different accelerometer devices can differ substantially in sampling frequency, data processing algorithms, and other characteristics [77]. Therefore, examining various accelerometers and possibly creating formulas for output transformation might be valuable to facilitate the comparison of outputs from different devices. In this study, we assessed 2 samples with currently depressed and nondepressed individuals with moderate to severe levels of depressive symptoms. Therefore, it remains to be seen whether the results of the constructed models generalize to a population with more ambiguous depressive symptoms or other psychiatric problems. A large-scale study in the general population is needed before recommendations can be made for the use of such prediction models in clinical practice. If this is the case, implementation of such a prediction tool in practice can be relatively straightforward. First, ESM seems to be a feasible tool for clinical practice and has the additional benefit that clients become co-owners of the care process [78,79]. Second, the developed algorithm for calculating “depression-related affect and behavior” sum scores is easy to use. However, further automation processes are needed to facilitate use of such a tool in real-life settings, such as primary care. Finally, such studies, including the current one, are necessary to build up empirical evidence on mechanisms, which will be the basis for developing medical devices in the future. Using strict regulations for such kinds of basic research might strangle innovative studies rather than pushing them forward.

Conclusions

To conclude, in this study we presented models to predict depression based on ESM-assessed depression-related affect and behavior as well as actigraphy data. The most potent predictor was the “depression-related affect and behavior” sum score constructed from the ESM data. The ESM model had an excellent predictive capacity, is easy to calculate, and hence might be feasible for future implementation in clinical practice. The actigraphy model had reasonable performance, but there was no added value of actigraphy in combination with ESM. Despite the fact that the actigraphy model had a lower predictive capacity than ESM, actigraphy could still be of value in situations where ESM is too burdensome.

Authors’ Contributors

OM, SHB, MW, and HR were involved in the design of the study. OM and SHB performed the statistical analysis. OM wrote the original draft of the manuscript. OM, SHB, MW, HR, FL, and NA critically revised the draft for intellectual content and approved the final version.

Acknowledgments

The infrastructure for NESDA is funded through the Geestkracht program of the Netherlands Organization for Health Research and Development (ZonMw, grant number 10-000-1002) and financial contributions by participating universities and mental health care organizations (VU University Medical Center, GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum). The authors thank all NESDA and MOOVD participants, research assistants, and students who made the data acquisition possible.

This study was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovative programme (ERC-CoG-2015; No 681466 to M Wichers).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart of a study sample from the Netherland Study of Depression and Anxiety-Ecological Momentary Assessment and Actigraphy (NESDA-EMAA) study.
[DOCX File , 32 KB - jmir_v22i12e22634_app1.docx]

Multimedia Appendix 2
List of included experience sampling method (ESM) items from the Netherland Study of Depression and Anxiety (NESDA) and the Mood and Movement in Daily Life (MOOVD) data sets and corresponding Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DMS-5) criteria.
[DOCX File , 17 KB - jmir_v22i12e22634_app2.docx]

Multimedia Appendix 3
Calibration plots of the models in the development data set.
[DOCX File , 86 KB - jmir_v22i12e22634_app3.docx]

Multimedia Appendix 4
The receiver operating characteristic (ROC) curves in the validation dataset.
[DOCX File , 35 KB - jmir_v22i12e22634_app4.docx]

Multimedia Appendix 5
Calibration plots of the models in the validation data set.
[DOCX File , 80 KB - jmir_v22i12e22634_app5.docx]

Multimedia Appendix 6
Sensitivity, specificity, and cutoff score for the experience sampling method (ESM) model, the actigraphy model, and the final (combined-domains) model in the development data set.
[DOCX File , 16 KB - jmir_v22i12e22634_app6.docx]

Multimedia Appendix 7
Sensitivity, specificity, and cutoff score for the experience sampling method (ESM) model, the actigraphy model, and the final (combined-domains) model in the validation data set.
[DOCX File , 13 KB - jmir_v22i12e22634_app7.docx]

Multimedia Appendix 8
Quantitative comparisons of the articles referred to in the discussion.
[DOCX File , 21 KB - jmir_v22i12e22634_app8.docx]

References

1. GBD 2017 DiseaseInjury IncidencePrevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Depression. World Health Organization. 2019. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2020-01-17]
3. van Weel C, van Weel-Baumgarten E, van Rijswijk E. Treatment of depression in primary care. BMJ 2009 Mar 19;338:b934. [doi: 10.1136/bmj.b934] [Medline: 19299476]
4. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 2009 Aug 22;374(9690):609-619. [doi: 10.1016/S0140-6736(09)60879-5] [Medline: 19640579]
5. Fernández A, Pinto-Meza A, Bellón JA, Roura-Poch P, Haro JM, Autonell J, et al. Is major depression adequately diagnosed and treated by general practitioners? Results from an epidemiological study. Gen Hosp Psychiatry 2010;32(2):201-209. [doi: 10.1016/j.genhosppsych.2009.11.015] [Medline: 20302995]

6. Carey M, Jones K, Meadows G, Sanson-Fisher R, D'Este C, Inder K, et al. Accuracy of general practitioner unassisted detection of depression. Aust N Z J Psychiatry 2014 Jan 10;48(6):571-578 [FREE Full text] [doi: 10.1177/0004867413520047] [Medline: 24413807]
7. Thornicroft G. No time to lose: onset and treatment delay for mental disorders. Epidemiol Psychiatr Sci 2012 Mar;21(1):59-61. [doi: 10.1017/s2045796011000825] [Medline: 22670413]
8. Wenze SJ, Miller IW. Use of ecological momentary assessment in mood disorders research. Clin Psychol Rev 2010 Aug;30(6):794-804. [doi: 10.1016/j.cpr.2010.06.007] [Medline: 20619520]
9. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
10. aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies. Clin Psychol Rev 2012 Aug;32(6):510-523. [doi: 10.1016/j.cpr.2012.05.007] [Medline: 22721999]
11. Armey MF, Schatten HT, Haradhvala N, Miller IW. Ecological momentary assessment (EMA) of depression-related phenomena. Curr Opin Psychol 2015 Aug 1;4:21-25. [doi: 10.1016/j.copsyc.2015.01.002] [Medline: 25664334]
12. Telford C, McCarthy-Jones S, Corcoran R, Rowse G. Experience sampling methodology studies of depression: the state of the art. Psychol Med 2012 Jun;42(6):1119-1129. [doi: 10.1017/S0033291711002200] [Medline: 22008511]
13. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
14. van de Water ATM, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography - a systematic review. J Sleep Res 2011 Mar;20(1 Pt 2):183-200 [FREE Full text] [doi: 10.1111/j.1365-2869.2009.00814.x] [Medline: 20374444]
15. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. Front Nutr 2014;1:5 [FREE Full text] [doi: 10.3389/fnut.2014.00005] [Medline: 25988109]
16. Kooiman TJM, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
17. Germain A, Kupfer DJ. Circadian rhythm disturbances in depression. Hum Psychopharmacol 2008 Oct;23(7):571-585 [FREE Full text] [doi: 10.1002/hup.964] [Medline: 18680211]
18. Hickie IB, Naismith SL, Robillard R, Scott EM, Hermens DF. Manipulating the sleep-wake cycle and circadian rhythms to improve clinical management of major depression. BMC Med 2013 Mar 22;11:79 [FREE Full text] [doi: 10.1186/1741-7015-11-79] [Medline: 23521808]
19. Vadnie CA, McClung CA. Circadian rhythm disturbances in mood disorders: insights into the role of the suprachiasmatic nucleus. Neural Plast 2017;2017:1504507 [FREE Full text] [doi: 10.1155/2017/1504507] [Medline: 29230328]
20. Burton C, McKinstry B, Szentagotai TA, Serrano-Blanco A, Pagliari C, Wolters M. Activity monitoring in patients with depression: a systematic review. J Affect Disord 2013 Feb 15;145(1):21-28. [doi: 10.1016/j.jad.2012.07.001] [Medline: 22868056]
21. Hori H, Koga N, Hidese S, Nagashima A, Kim Y, Higuchi T, et al. 24-h activity rhythm and sleep in depressed outpatients. J Psychiatr Res 2016 Jun;77:27-34. [doi: 10.1016/j.jpsychires.2016.02.022] [Medline: 26978182]
22. Tazawa Y, Wada M, Mitsukura Y, Takamiya A, Kitazawa M, Yoshimura M, et al. Actigraphy for evaluation of mood disorders: A systematic review and meta-analysis. J Affect Disord 2019 Apr 22;253:257-269. [doi: 10.1016/j.jad.2019.04.087] [Medline: 31060012]
23. Tonon AC, Fuchs DFP, Barbosa Gomes W, Levandovski R, Pio de Almeida Fleck M, Hidalgo MPL, et al. Nocturnal motor activity and light exposure: Objective actigraphy-based marks of melancholic and non-melancholic depressive disorder. Brief report. Psychiatry Res 2017 Dec;258:587-590. [doi: 10.1016/j.psychres.2017.08.025] [Medline: 28844556]
24. Paudel ML, Taylor BC, Ancoli-Israel S, Stone KL, Tranah G, Redline S, et al. Rest/activity rhythms and cardiovascular disease in older men. Chronobiol Int 2011 Apr;28(3):258-266 [FREE Full text] [doi: 10.3109/07420528.2011.553016] [Medline: 21452921]
25. Laposky AD, Bass J, Kohsaka A, Turek FW. Sleep and circadian rhythms: key components in the regulation of energy metabolism. FEBS Lett 2008 Jan 09;582(1):142-151 [FREE Full text] [doi: 10.1016/j.febslet.2007.06.079] [Medline: 17707819]
26. Atkinson G, Davenne D. Relationships between sleep, physical activity and human health. Physiol Behav 2007 Feb 28;90(2-3):229-235 [FREE Full text] [doi: 10.1016/j.physbeh.2006.09.015] [Medline: 17067643]
27. Quante M, Mariani S, Weng J, Marinac CR, Kaplan ER, Rueschman M, et al. Zeitgebers and their association with rest-activity patterns. Chronobiol Int 2019 Feb;36(2):203-213. [doi: 10.1080/07420528.2018.1527347] [Medline: 30365354]
28. Finazzi ME, Mesquita ME, Lopes JR, Fu LI, Oliveira MG, Del Porto JA. Motor activity and depression severity in adolescent outpatients. Neuropsychobiology 2010;61(1):33-40. [doi: 10.1159/000262178] [Medline: 19940518]

29. Luik AI, Zuurbier LA, Hofman A, van Someren EJW, Tiemeier H. Stability and fragmentation of the activity rhythm across the sleep-wake cycle: the importance of age, lifestyle, and mental health. Chronobiol Int 2013 Dec;30(10):1223-1230. [doi: 10.3109/07420528.2013.813528] [Medline: 23971909]
30. Maglione JE, Ancoli-Israel S, Peters KW, Paudel ML, Yaffe K, Ensrud KE, Study of Osteoporotic Fractures Research Group. Depressive symptoms and circadian activity rhythm disturbances in community-dwelling older women. Am J Geriatr Psychiatry 2014 Apr;22(4):349-361 [FREE Full text] [doi: 10.1016/j.jagp.2012.09.003] [Medline: 23567424]
31. Difrancesco S, Lamers F, Riese H, Merikangas KR, Beekman ATF, van Hemert AM, et al. Sleep, circadian rhythm, and physical activity patterns in depressive and anxiety disorders: A 2-week ambulatory assessment study. Depress Anxiety 2019 Oct;36(10):975-986 [FREE Full text] [doi: 10.1002/da.22949] [Medline: 31348850]
32. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression prediction by using Ecological Momentary Assessment, actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14149 [FREE Full text] [doi: 10.2196/14149] [Medline: 31621642]
33. Penninx BWJH, Beekman ATF, Smit JH, Zitman FG, Nolen WA, Spinhoven P, NESDA Research Consortium. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int J Methods Psychiatr Res 2008;17(3):121-140 [FREE Full text] [doi: 10.1002/mpr.256] [Medline: 18763692]
34. Booij SH, Bos EH, Bouwmans MEJ, van Faassen M, Kema IP, Oldehinkel AJ, et al. Cortisol and α-amylase secretion patterns between and within depressed and non-depressed individuals. PLoS One 2015 Jul 6;10(7):e0131002 [FREE Full text] [doi: 10.1371/journal.pone.0131002] [Medline: 26148294]
35. Schoevers RA, van Borkulo CD, Lamers F, Servaas M, Bastiaansen JA, Beekman ATF, et al. Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. Psychol Med 2020 Apr 01:1-10. [doi: 10.1017/S0033291720000689] [Medline: 32234092]
36. Wittchen HU. Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res 1994;28(1):57-84. [doi: 10.1016/0022-3956(94)90036-1] [Medline: 8064641]
37. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 1996 May;26(3):477-486. [doi: 10.1017/s0033291700035558] [Medline: 8733206]
38. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. San Antonio: TX: Psychological Corporation; 1996:1-82.
39. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. Med Sci Sports Exerc 2011 Jun;43(6):1085-1093. [doi: 10.1249/MSS.0b013e31820513be] [Medline: 21088628]
40. Pavey TG, Gomersall SR, Clark BK, Brown WJ. The validity of the GENEActiv wrist-worn accelerometer for measuring adult sedentary time in free living. J Sci Med Sport 2016 May;19(5):395-399. [doi: 10.1016/j.jsams.2015.04.007] [Medline: 25956687]
41. Hager ER, Treuth MS, Gormely C, Epps L, Snitker S, Black MM. Ankle accelerometry for assessing physical activity among adolescent girls: threshold determination, validity, reliability, and feasibility. Res Q Exerc Sport 2015;86(4):397-405 [FREE Full text] [doi: 10.1080/02701367.2015.1063574] [Medline: 26288333]
42. Stavrakakis N, Booij SH, Roest AM, de Jonge P, Oldehinkel AJ, Bos EH. Temporal dynamics of physical activity and affect in depressed and nondepressed individuals. Health Psychol 2015 Dec;34S:1268-1277. [doi: 10.1037/hea0000303] [Medline: 26651468]
43. Bouwmans MEJ, Bos EH, Booij SH, van Faassen M, Oldehinkel AJ, de Jonge P. Intra- and inter-individual variability of longitudinal daytime melatonin secretion patterns in depressed and non-depressed individuals. Chronobiol Int 2015 Apr 27;32(3):441-446. [doi: 10.3109/07420528.2014.973114] [Medline: 25347155]
44. American Psychiatric Association: Diagnostic and statistical manual of mental disorders, fourth edition, text revision. Washington, DC: American Psychiatric Association; 2000.
45. Vallance JK, Winkler EA, Gardiner PA, Healy GN, Lynch BM, Owen N. Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005-2006). Prev Med 2011 Oct;53(4-5):284-288. [doi: 10.1016/j.ypmed.2011.07.013] [Medline: 21820466]
46. Robillard R, Naismith SL, Smith KL, Rogers NL, White D, Terpening Z, et al. Sleep-wake cycle in young and older persons with a lifetime history of mood disorders. PLoS One 2014;9(2):e87763 [FREE Full text] [doi: 10.1371/journal.pone.0087763] [Medline: 24586290]
47. Sadeh A, Alster J, Urbach D, Lavie P. Actigraphically based automatic bedtime sleep-wake scoring: Validity and clinical applications. J Ambul Monit 1989;2(3):209-216.
48. Marler MR, Gehrman P, Martin JL, Ancoli-Israel S. The sigmoidally transformed cosine curve: a mathematical model for circadian rhythms with symmetric non-sinusoidal shapes. Stat Med 2006 Nov 30;25(22):3893-3904. [doi: 10.1002/sim.2466] [Medline: 16381069]
49. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2012 May 18;36(1):27-46. [doi: 10.1111/j.1600-0587.2012.07348.x]
50. Field A. Discovering Statistics Using IBM SPSS Statistics. Los Angeles, CA: Sage Publications Ltd; 2013.

51. Piccinelli M, Wilkinson G. Gender differences in depression. Critical review. Br J Psychiatry 2000 Dec;177:486-492. [doi: 10.1192/bjp.177.6.486] [Medline: 11102321]
52. Miech RA, Shanahan MJ. Socioeconomic status and depression over the life course. J Health Soc Behav 2000 Jun;41(2):162-176. [doi: 10.2307/2676303]
53. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med 2000 Apr 30;19(8):1059-1079. [doi: 10.1002/(sici)1097-0258(20000430)19:8<1059::aid-sim412>3.0.co;2-0] [Medline: 10790680]
54. Harrell FE. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis. Springer International Publishing AG Switzerland: Springer Ser Stat; 2015.
55. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010 Jan;21(1):128-138 [FREE Full text] [doi: 10.1097/EDE.0b013e3181c30fb2] [Medline: 20010215]
56. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015 Jan 06;162(1):W1-73 [FREE Full text] [doi: 10.7326/M14-0698] [Medline: 25560730]
57. Kim J, Nakamura T, Yamamoto Y. A momentary biomarker for depressive mood. In Silico Pharmacol 2016 Dec;4(1):4 [FREE Full text] [doi: 10.1186/s40203-016-0017-6] [Medline: 26979449]
58. de Jonge M, Dekker JJM, Kikkert MJ, Peen J, van Rijsbergen GD, Bockting CLH. The role of affect in predicting depressive symptomatology in remitted recurrently depressed patients. J Affect Disord 2017 Mar 01;210:66-71. [doi: 10.1016/j.jad.2016.12.015] [Medline: 28013124]
59. Wichers M, Peeters F, Geschwind N, Jacobs N, Simons CJP, Derom C, et al. Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: a momentary assessment study. J Affect Disord 2010 Jul;124(1-2):191-195. [doi: 10.1016/j.jad.2009.11.010] [Medline: 20004977]
60. Wichers M, Lothmann C, Simons CJP, Nicolson NA, Peeters F. The dynamic interplay between negative and positive emotions in daily life predicts response to treatment in depression: a momentary assessment study. Br J Clin Psychol 2012 Jun;51(2):206-222. [doi: 10.1111/j.2044-8260.2011.02021.x] [Medline: 22574805]
61. Geschwind N, Nicolson NA, Peeters F, van Os J, Barge-Schaapveld D, Wichers M. Early improvement in positive rather than negative emotion predicts remission from depression after pharmacotherapy. Eur Neuropsychopharmacol 2011 Mar;21(3):241-247 [FREE Full text] [doi: 10.1016/j.euroneuro.2010.11.004] [Medline: 21146375]
62. Cohen JR, Thakur H, Young JF, Hankin BL. The development and validation of an algorithm to predict future depression onset in unselected youth. Psychol Med 2019 Oct 02:1-9. [doi: 10.1017/S0033291719002691] [Medline: 31576786]
63. Andrews JA, Harrison RF, Brown LJE, MacLean LM, Hwang F, Smith T, et al. Using the NANA toolkit at home to predict older adults' future depression. J Affect Disord 2017 Apr 15;213:187-190 [FREE Full text] [doi: 10.1016/j.jad.2017.02.019] [Medline: 28259086]
64. Pyszczynski T, Hamilton JC, Herring FH, Greenberg J. Depression, self-focused attention, and the negative memory bias. J Pers Soc Psychol 1989 Mar;57(2):351-357 [FREE Full text] [doi: 10.1037/0022-3514.57.2.351]
65. Karas M, Bai J, Strączkiewicz M, Harezlak J, Glynn NW, Harris T, et al. Accelerometry data in health research: challenges and opportunities. Stat Biosci 2019 Jul;11(2):210-237 [FREE Full text] [doi: 10.1007/s12561-018-9227-2] [Medline: 31762829]
66. Reichert M, Lutz A, Deuschle M, Gilles M, Hill H, Limberger MF, et al. Improving motor activity assessment in depression: which sensor placement, analytic strategy and diurnal time frame are most powerful in distinguishing patients from controls and monitoring treatment effects. PLoS One 2015;10(4):e0124231 [FREE Full text] [doi: 10.1371/journal.pone.0124231] [Medline: 25885258]
67. Todder D, Caliskan S, Baune BT. Longitudinal changes of day-time and night-time gross motor activity in clinical responders and non-responders of major depression. World J Biol Psychiatry 2009;10(4):276-284. [doi: 10.3109/15622970701403081] [Medline: 19921969]
68. Smagula SF, Boudreau RM, Stone K, Reynolds CF, Bromberger JT, Ancoli-Israel S, Osteoporotic Fractures in Men (MrOS) Research Group. Latent activity rhythm disturbance sub-groups and longitudinal change in depression symptoms among older men. Chronobiol Int 2015;32(10):1427-1437 [FREE Full text] [doi: 10.3109/07420528.2015.1102925] [Medline: 26594893]
69. McClung CA. Circadian rhythms in mood disorders. In: Circadian Medicine. Hoboken, NJ: John Wiley & Sons, Inc; 2015.
70. McClung CA. How might circadian rhythms control mood? Let me count the ways. Biol Psychiatry 2013 Aug 15;74(4):242-249 [FREE Full text] [doi: 10.1016/j.biopsych.2013.02.019] [Medline: 23558300]
71. Lyall LM, Wyse CA, Graham N, Ferguson A, Lyall DM, Cullen B, et al. Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. Lancet Psychiatry 2018 Jun;5(6):507-514. [doi: 10.1016/S2215-0366(18)30139-1] [Medline: 29776774]
72. Nebes RD, Buysse DJ, Halligan EM, Houck PR, Monk TH. Self-reported sleep quality predicts poor cognitive performance in healthy older adults. J Gerontol B Psychol Sci Soc Sci 2009 Mar;64(2):180-187 [FREE Full text] [doi: 10.1093/geronb/gbn037] [Medline: 19204069]

73. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol 2008 Nov;61(11):1085-1094. [doi: 10.1016/j.jclinepi.2008.04.008] [Medline: 19208371]
74. Mata J, Thompson RJ, Jaeggi SM, Buschkuehl M, Jonides J, Gotlib IH. Walk on the bright side: physical activity and affect in major depressive disorder. J Abnorm Psychol 2012 May;121(2):297-308 [FREE Full text] [doi: 10.1037/a0023533] [Medline: 21553939]
75. Thompson RJ, Mata J, Jaeggi SM, Buschkuehl M, Jonides J, Gotlib IH. The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. J Abnorm Psychol 2012 Nov;121(4):819-829 [FREE Full text] [doi: 10.1037/a0027978] [Medline: 22708886]
76. Bylsma LM, Taylor-Clift A, Rottenberg J. Emotional reactivity to daily events in major and minor depression. J Abnorm Psychol 2011 Feb;120(1):155-167. [doi: 10.1037/a0021662] [Medline: 21319928]
77. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle NC, Mora-Gonzalez J, Löf M, et al. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. Sports Med 2017 Sep;47(9):1821-1845. [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
78. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, ESM-MERGE Investigators PhD, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 Dec;34(6):481-493. [doi: 10.1002/da.22647] [Medline: 28544391]
79. Bos FM, Snippe E, Bruggeman R, Wichers M, van der Krieke L. Insights of patients and clinicians on the promise of the experience sampling method for psychiatric care. Psychiatr Serv 2019 Nov 01;70(11):983-991. [doi: 10.1176/appi.ps.201900050] [Medline: 31434558]

Abbreviations

AC: activity counts
AIC: Akaike information criterion
AUC: the area under the receiver operating characteristic curve
BDI-II: Beck Depression Inventory-II
CIDI: Composite International Diagnostic Interview
DSM: Diagnostic and Statistical Manual of Mental Disorders
ENMO: Euclidian Norm Minus One ESM: experience sampling method
IDS-SR: Inventory of Depressive Symptomatology, self-report
MESOR: midline estimating statistic of rhythm
MOOVD: Mood and Movement in Daily Life
MVPA: moderate-to-vigorous physical activity
NESDA: the Netherlands Study of Depression and Anxiety
NESDA-EMAA: Ecological Momentary Assessment and Actigraphy substudy of NESDA
RAR: rest-activity rhythm
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
TST: total sleep time
VIF: variance inflation factor

Edited by G Eysenbach; submitted 19.07.20; peer-reviewed by J Li, Y Li; comments to author 28.07.20; revised version received 13.08.20; accepted 26.10.20; published 01.12.20.

Please cite as:
Minaeva O, Riese H, Lamers F, Antypa N, Wichers M, Booij SH
Screening for Depression in Daily Life: Development and External Validation of a Prediction Model Based on Actigraphy and Experience Sampling Method
J Med Internet Res 2020;22(12):e22634
URL: https://www.jmir.org/2020/12/e22634
doi:10.2196/22634
PMID:33258783

Original Paper

A Web-Based Mental Health Intervention to Improve Social and Occupational Functioning in Adults With Type 2 Diabetes (The Springboard Trial): 12-Month Outcomes of a Randomized Controlled Trial

Peter Andrew Baldwin[1], PhD; Samineh Sanatkar[1], PhD; Janine Clarke[1], PhD; Susan Fletcher[2], PhD; Jane Gunn[2], MD, PhD; Kay Wilhelm[3], MBBS, MD; Lesley Campbell[3], MBBS, FRCP, FRACP, AM; Nicholas Zwar[3], MBBS, PhD; Mark Harris[3], MBBS, MD; Helen Lapsley[3], PhD; Dusan Hadzi-Pavlovic[1], BSc, MPsychol; Helen Christensen[1], PhD; Judy Proudfoot[1], PhD

[1]Black Dog Institute, Randwick, Australia
[2]University of Melbourne, Melbourne, Australia
[3]UNSW Sydney, Randwick, Australia

Corresponding Author:
Peter Andrew Baldwin, PhD
Black Dog Institute
Hospital Road
Randwick, 2031
Australia
Phone: 61 0293824521
Email: p.baldwin@unsw.edu.au

Abstract

Background: People with type 2 diabetes mellitus (T2DM) often experience mental health symptoms that exacerbate illness and increase mortality risk. Access to psychological support is low in people with T2DM. Detection of depression is variable in primary care and can be further hampered by mental health stigma. Electronic mental health (eMH) programs may provide an accessible, private, nonstigmatizing mental health solution for this group.

Objective: This study aims to evaluate the efficacy over 12 months of follow-up of an eMH program (myCompass) for improving social and occupational functioning in a community sample of people with T2DM and self-reported mild-to-moderate depressive symptoms. myCompass is a fully automated and self-guided web-based public health program for people with depression or anxiety. The effects of myCompass on depressive symptoms, diabetes-related distress, anxiety symptoms, and self-care behavior were also examined.

Methods: Adults with T2DM and mild-to-moderate depressive symptoms (N=780) were recruited via online advertisements, community organizations, and general practices. Screening, consent, and self-report questionnaires were administered online. Eligible participants were randomized to receive either myCompass (n=391) or an attention control generic health literacy program (Healthy Lifestyles; n=379) for 8 weeks. At baseline and at 3, 6, and 12 months postintervention, participants completed the Work and Social Adjustment Scale, the Patient Health Questionnaire-9 item, the Diabetes Distress Scale, the Generalized Anxiety Disorder Questionnaire-7 item, and items from the Self-Management Profile for Type 2 Diabetes. Glycosylated hemoglobin measurements were obtained at baseline and 6 and 12 months postintervention.

Results: A total of 38.9% (304/780) of the trial participants completed all postintervention assessments. myCompass users logged in on an average of 6 times and completed an average of 0.29 modules. Healthy Lifestyles users logged in on an average of 4 times and completed an average of 1.37 modules. At baseline, the mean scores on several outcome measures, including the primary outcome of work and social functioning, were close to the normal range, despite a varied and extensive recruitment process. Intention-to-treat analyses revealed slightly greater improvement at 12 months in work and social functioning for the Healthy Lifestyles group relative to the myCompass group. All participants reported equivalent improvements in depression anxiety, diabetes distress, diabetes self-management, and glycemic control across the trial.

Conclusions: The Healthy Lifestyles group reported higher ratings of social and occupational functioning than the myCompass group, but no differences were observed for any secondary outcome. Although these findings should be interpreted in light of the near-floor symptom scores at baseline, the trial yields important insights into how people with T2DM might be engaged in eMH programs and the challenges of focusing specifically on mental health. Several avenues emerge for continued investigation into how best to deal with the growing mental health burden in adults with T2DM.

Trial Registration: Australian New Zealand Clinical Trials Registry Number (ACTRN) 12615000931572; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=368109&isReview=true

(J Med Internet Res 2020;22(12):e16729) doi:10.2196/16729

KEYWORDS

type 2 diabetes; depression; internet

Introduction

Background

Type 2 diabetes mellitus (T2DM) affects over 1 million Australians [1] and increases the risk of psychiatric and neurodegenerative disorders [2,3]. Up to 40% of people with T2DM experience depressive symptoms [4,5], which appear to worsen the physical health via impaired psychosocial functioning, poorer self-care, and increased need for outpatient and inpatient health services [2,4,5]. The relationship between T2DM and depressive symptoms seems to be bidirectional [6], with one condition intensifying symptoms in the other [4]. Therefore, population-based mental health programs for depression in people with T2DM have the potential to reduce the substantial personal burden and public health impacts of these comorbid conditions.

Cognitive behavioral therapy (CBT) is the most established psychological treatment for depression. In people with T2DM, face-to-face CBT appears to be effective in improving not only depressive symptoms but also fasting glucose levels [7], self-care behaviors, and overall quality of life [8]. Although most T2DM management takes place in primary care, detecting depression in people with T2DM can be challenging, and it may not be feasible for general practitioners (GPs) to deal adequately with both mental health and diabetes care within a single consultation. Concerns about mental health stigma, treatment cost, and *clinician fatigue* can lead some patients to avoid seeking help [8]. The high rates of comorbid T2DM and depressive symptoms [9] also mean that already-stretched local health systems may struggle to provide services to every person needing help [10,11]. Scalable methods of delivering evidence-based psychological therapies may provide an answer to many of these challenges [7].

Electronic mental health (eMH) programs can be clinically effective and cost-efficient tools for increasing the availability of mental health services [12,13]. In previous trials, eMH programs have been effective in addressing both depressive symptoms and diabetes-related distress in people with T2DM, using both diabetes-specific content [14] or existing depression treatments that incorporate therapist assistance [15]. Importantly, data suggest that eMH interventions are most effective in the mild-to-moderate depressive symptom range [16] prevalent in T2DM [17,18]. Therefore, low-intensity, population-based eMH programs seem well suited to people with T2DM who are also experiencing problems with low mood.

myCompass is an eMH program for depression and anxiety, which, in contrast to the eMH programs described earlier, operates as a self-help program without therapist assistance or diabetes content. A large randomized controlled trial (RCT) demonstrated that myCompass users with mild-to-moderate depression and anxiety experienced a significant improvement in symptoms and functioning, compared with placebo [19]. Data from a feasibility study indicated that myCompass has the potential to have a positive impact on the functioning and depressive symptoms in people with diabetes [20]. Therefore, the primary aim of the SpringboarD trial was to evaluate the impact of myCompass on work and social functioning in adults with T2DM and mild-to-moderate depressive symptoms relative to a generic health literacy program. In our primary analysis at 3-month postintervention, both groups reported significant improvements, but there was no specific benefit of myCompass [21].

Objectives

Further to our analysis of 3-month outcomes [21], the aim of this study was to establish if allocation to the myCompass intervention resulted in improvement of daily functioning of adults with T2DM and mild-to-moderate depressive symptoms across a 12-month period. We hypothesized that the participants using myCompass would report improvements in self-reported work and social functioning relative to the participants using a placebo health literacy program at the 6-month and 12-month follow-up. In addition to examining long-term outcomes from our trial (RCT), this study also explored long-term changes in health and clinical outcomes across the 12-month period of the trial and included a biological marker of glycemic control (glycosylated hemoglobin, HbA_{1c}). Inclusion of a wide range of variables in this study enabled us to identify which psychosocial factors most affected change in both physical and mental health outcomes across the trial.

Methods

Design

This paper is a secondary analysis of a two-arm RCT called *SpringboarD*. The full SpringboarD trial protocol is detailed elsewhere [22]. Across the trial, outcomes were assessed at baseline and at 3-, 6-, and 12-month postrandomization. All the participants had uninterrupted access to usual diabetes treatment throughout the study. The study was approved by the Human Research Ethics Committee (HREC) at UNSW Sydney (HREC 15090) and registered with the Australia and New Zealand Clinical Trials Register (ACTRN12615000931572).

Participants and Setting

Full details of recruitment for the SpringboarD trial are detailed in separate papers [21,23]. In summary, recruitment began in September 2015 and continued until November 2017. The trial was advertised via GPs in New South Wales and Victoria, professional associations (eg, the Australian Association of Practice Managers), print advertisements in national diabetes-related publications, and online (via Google and Facebook). Potential candidates were contacted via email through Black Dog Institute's Volunteer Research Register and the Sax Institute's *45 and Up Study*, a large, longitudinal cohort study of healthy aging described elsewhere [24]. All promotional materials directed potential candidates to a secure study-specific website, which guided interested participants through the consent process and provided instructions regarding completion of the screening questionnaires.

Eligibility Criteria

Australian residents were eligible for SpringboarD if they were aged 18-75 years, diagnosed with T2DM by a medical doctor, scored ≥2 on the 2-item Patient Health Questionnaire [25] (indicating likely depression), and had access to an internet-connected device. People who scored ≥2 on the 2-item Patient Health Questionnaire proceeded to complete the Patient Health Questionnaire-9 (PHQ-9) [26] at screening to establish depressive symptom levels. People were excluded if they answered *no* to the question, *Are you able to read and write English easily?*; had extremely severe depressive symptoms on the full PHQ-9 (score > 19); had probable psychosis (measured by the psychosis screener developed for the Australian National Mental Health and Well-being Survey) [27]; were currently receiving face-to-face therapy for depression; had a recent (within 2 months) change in antidepressant medication; had an elevated suicide risk (assessed by item 9 of the PHQ-9); or had used the myCompass program previously. Participants ineligible due to severe depressive symptoms, elevated suicide risk, or probable psychosis were referred to professional mental health services.

Randomization

Participants were allocated to the intervention and control conditions using computerized block randomization at a 1:1 allocation ratio, which was initiated automatically by the Black Dog Institute study management software after the completion of the baseline questionnaires. Allocation was concealed from participants and researchers.

Interventions

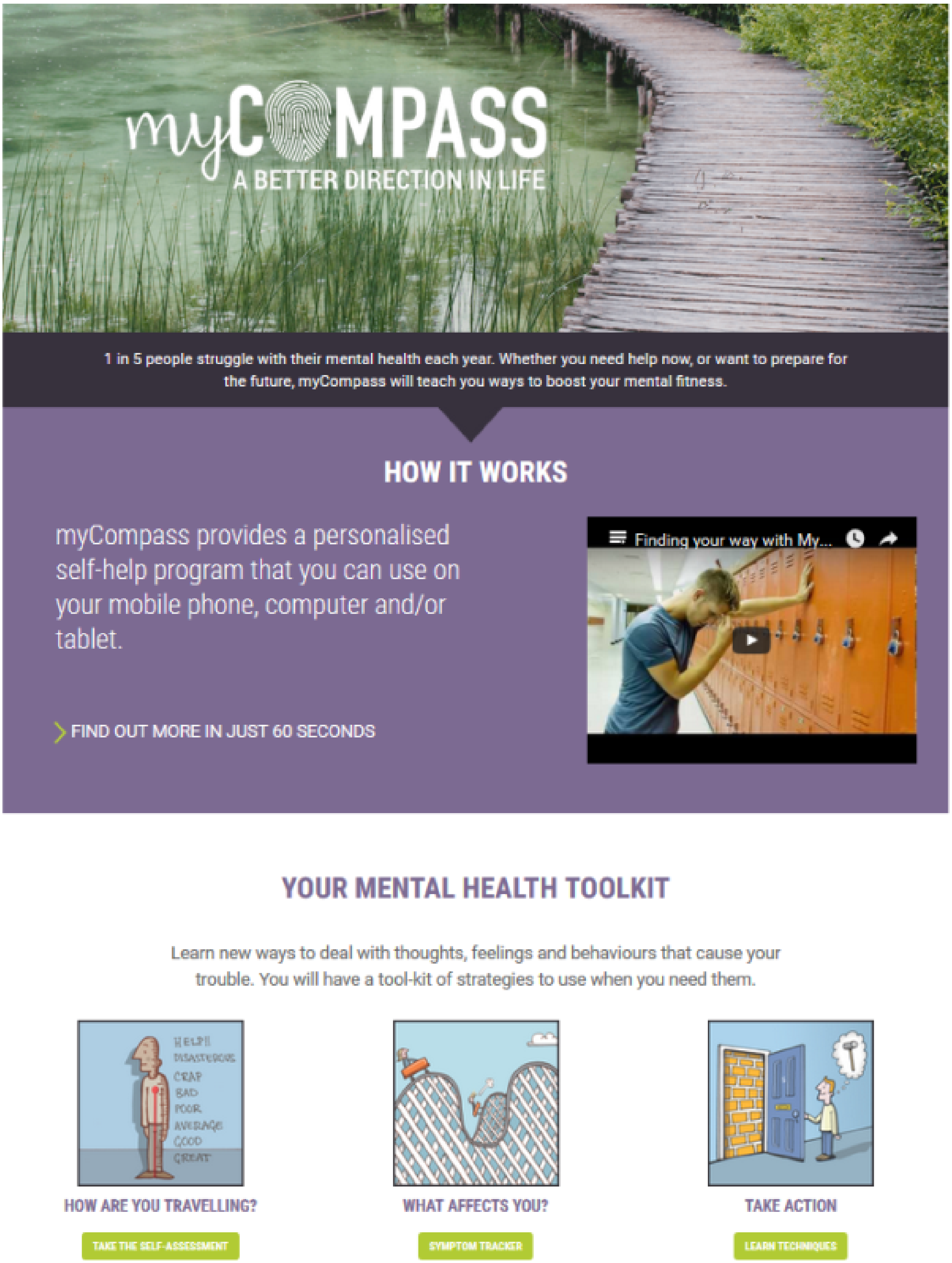
Active Intervention (myCompass)

myCompass (Figure 1) is a fully automated eMH program [19] that contains 12 interactive mental health modules and allows users to self-monitor 3 of the total of 20 cognitive behavioral variables, such as eating, mood, or anxiety (Figure 2). Users can freely select modules and self-monitoring variables or opt for algorithm-based online guidance based on self-reported mental health symptoms given at registration or during program use. myCompass also provides SMS and/or email reminders, home practice activities, mental health care tips, motivational statements, and graphical reporting of self-monitoring data.

Participants randomized to the myCompass condition had full program access for 8 weeks, followed by a 4-week tailing-off period in which only the self-monitoring function was available. The program recommended that users complete 3 mental health modules and self-monitor up to 3 cognitive behavioral variables daily. myCompass users also received automated feedback via email about their program use in weeks 1, 3, 5, and 7.

myCompass user privacy is maintained via a password-protected login. All data are encrypted during transmission and stored on secure servers rather than on the user's device. Participants' myCompass user data were identified using the email address provided during the study registration. Once extracted from the myCompass server, data were de-identified and stored in a password-protected file.

Figure 1. Screenshot of the myCompass landing page.



1 in 5 people struggle with their mental health each year. Whether you need help now, or want to prepare for the future, myCompass will teach you ways to boost your mental fitness.

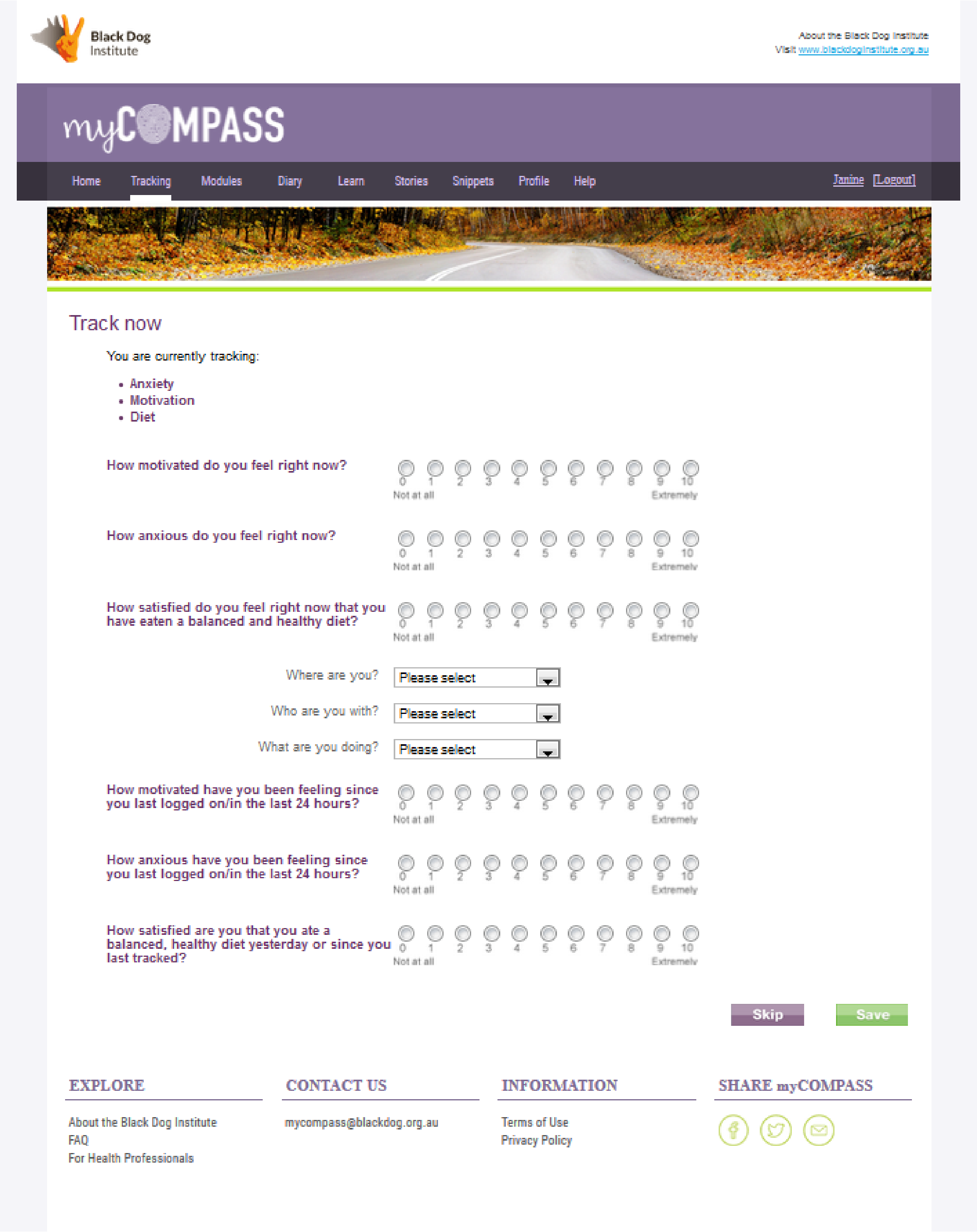
HOW IT WORKS

myCompass provides a personalised self-help program that you can use on your mobile phone, computer and/or tablet.

YOUR MENTAL HEALTH TOOLKIT

Learn new ways to deal with thoughts, feelings and behaviours that cause your trouble. You will have a tool-kit of strategies to use when you need them.

Figure 2. Screenshot of the myCompass self-monitoring page.



Track now

You are currently tracking:

- Anxiety
- Motivation
- Diet

How motivated do you feel right now?
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10
Not at all … Extremely

How anxious do you feel right now?
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10
Not at all … Extremely

How satisfied do you feel right now that you have eaten a balanced and healthy diet?
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10
Not at all … Extremely

Where are you? Please select
Who are you with? Please select
What are you doing? Please select

How motivated have you been feeling since you last logged on/in the last 24 hours?
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10
Not at all … Extremely

How anxious have you been feeling since you last logged on/in the last 24 hours?
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10
Not at all … Extremely

How satisfied are you that you ate a balanced, healthy diet yesterday or since you last tracked?
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10
Not at all … Extremely

Attention Control Intervention (Healthy Lifestyles)

The attention control program, *Healthy Lifestyles*, was adapted from a previous attention control program [28] to replicate the online, module-based structure of myCompass, without delivering any active CBT content. The 12 Healthy Lifestyles modules provide lifestyle information and interactive activities across 8 topics: (1) eye care, (2) skin care, (3) mental health, (4) home safety and comfort, (5) healthy interactions with digital devices, (6) safe driving, (7) travel planning, and (8) healthy eating. Participants in the Healthy Lifestyles condition received login reminders via email at weeks 1, 3, 5, and 7, but they did not receive feedback regarding their program use. To reflect the SMS functionality of myCompass, Healthy Lifestyles users also received a weekly SMS containing general health and well-being information for the first 4 weeks of the intervention period. Participants had full access to the Healthy Lifestyles program for 8 weeks.

Outcome Measures

A summary of the measures obtained from participants at baseline and at 3-, 6-, and 12-month postrandomization is presented in Table 1.

Table 1. Measures obtained at each assessment phase.

Measures	Baseline	3 months	6 months	12 months
Demographic and disease-related information				
Demographic data	✓[a]	—[b]	—	—
Disease-relevant data	✓	✓	✓	✓
Mental health history	✓	—	—	—
Primary outcome				
WSAS[c]	✓	✓	✓	✓
Secondary outcomes				
PHQ-9[d]	✓	✓	✓	✓
DDS[e]	✓	✓	✓	✓
GAD-7[f]	✓	✓	✓	✓
SMP-T2D[g]	✓	✓	✓	✓
HbA_{1c}[h]	✓	—	✓	✓
Days out of role	✓	✓	✓	✓
Health service utilization	✓	✓	✓	✓

[a] Measurement taken.

[b] Measurement not taken.

[c] WSAS: Work and Social Adjustment Scale.

[d] PHQ-9: Patient Health Questionnaire-9.

[e] DDS: Diabetes Distress Scale.

[f] GAD-7: Generalized Anxiety Disorder Scale.

[g] SMP-T2D: Self-Management Profile for Type 2 Diabetes.

[h] HbA_{1c}: glycosylated hemoglobin.

Primary Outcome

The primary outcome for the trial was work and social functioning, measured by the Work and Social Adjustment Scale (WSAS). WSAS measures the daily functioning across 5 life domains: work, socializing, leisure, home, and personal relationships [29,30]. Scores range from 0 to 40, with higher scores indicating poorer functioning.

Secondary Outcomes

Depressive symptoms were measured using the PHQ-9 scale [20], and anxiety symptoms were measured using the 7-item Generalized Anxiety Disorder Scale (GAD-7) [31]. Both scales are well validated and commonly used as outcome measures in chronically ill cohorts [32,33]. Each scale uses cutoff scores of 5, 10, and 15 to reflect mild, moderate, and moderately severe symptoms, respectively.

Diabetes distress—a person’s emotional adjustment to diabetes—was measured using the Diabetes Distress Scale (DDS) [34]. DDS is a 17-item questionnaire that provides an overall measure of diabetes distress along with 4 subscale scores that index (1) regimen-related distress, (2) interpersonal distress, (3) emotional burden of diabetes, and (4) distress related to interacting with health care providers. DDS total and subscale scores were calculated by averaging all items in the scale or subscale and ranged from 1 to 6, with higher scores indicating greater distress. A score of >3 indicates clinically significant distress.

Diabetes management was measured using a subset of items from the Self-Management Profile for Type 2 Diabetes (SMP-T2D). The SMP-T2D was developed to measure the level and perceived ease of engaging in common diabetes management behaviors, along with perceived coping, confidence dealing with diabetes, and ease of weight management [35]. As DDS also asks about coping and confidence in diabetes self-care, to reduce participant assessment burden, we administered only the SMP-T2D items that assess diabetes management behaviors across 4 domains: blood glucose monitoring, medication adherence, healthy eating, and exercise. Scores in each behavior domain are converted to a percentage of the previous week spent engaging in a particular diabetes management behavior. Higher scores indicate more time spent on diabetes management [35].

Additional Measurements

We also collected baseline data regarding each participant’s diabetes history (eg, age of onset and treatment regimen), demographics (eg, age, gender, education, and occupation), and mental health history (eg, service use and previous diagnoses). With written consent, participants' most recent HbA_{1c} results were provided by their GPs at baseline and 6- and 12-month follow-up. Service utilization for physical and mental health was captured at baseline and all follow-up points, as was days out of role, defined as the number of days in the previous 30 that participants were unable to perform work or normal activities because of problems with physical or mental health [36].

Sample Size

Initial calculations identified that approximately 600 participants were required to detect a between-group effect of 0.3 in scores on the WSAS postintervention, with power of 80% and a two-tailed alpha level of .05, and a prediction of approximately 40% total attrition. Due to unexpectedly high attrition early in the trial, a further 180 participants were recruited to maintain sufficient power.

Analyses

All mixed model repeated measure (MMRM) analyses were computed in *R* [37] using the Linear Models module in Jamovi v0.9 [38]. Model parameters were estimated using restricted maximum likelihood estimation, and error degrees of freedom were calculated using the Satterthwaite approximation. Repeated measures (level 1) were nested within individual levels (level 2), and a random intercept was fitted at the individual level to account for intra-individual correlations on repeated measures.

Results

Data Preparation

All data were inspected for outliers, skew, and kurtosis. Data from the Medication Adherence and Exercise domains for the SMP-T2D were highly left-skewed and kurtotic. Therefore, these data were not included in subsequent analyses. All regression coefficients were reported in unstandardized format and, therefore, gave an estimate of the effect size in dependent-variable units. For example, our estimate of the effect of age on PHQ-9 scores over time was –0.089, indicating that each year of increased age was associated with a reduction in PHQ-9 of 0.089 points.

Sample Characteristics

Sample characteristics are presented in Table 2, and CONSORT (Consolidated Standards of Reporting Trials) metrics are shown in Figure 3. A total of 52.49% (3223/6145) of the individuals who visited the study website consented to online screening, and 27.55% (888/3223) of those participants were eligible to proceed to the baseline assessment. The most common reasons for ineligibility were the absence of depressive symptoms, that is, a score of <2 on the PHQ-2 (1021/1961, 52.1%), current face-to-face mental health treatment (516/1961, 26.31%), and severe depressive symptoms (183/1961, 9.3%). In total, 87.8% (780/888) of eligible individuals completed the baseline measures and were randomized and 7.3% (57/780) subsequently withdrew consent, leaving a final sample of 81.4% (723/888) of the eligible participants.

Table 2 presents the intervention and control group participants' characteristics at baseline. Participants were, on average, aged 58 years (SDs were 10.6 for intervention group participants and 10.0 for control group participants) and were mostly female (437/723, 60.4%). Randomization successfully matched the groups based on demographic characteristics as well as mental health– and diabetes-related histories. Participants allocated to the intervention group reported taking antidepressants for slightly longer (mean 98 days, SD 95 days) than the control group participants (mean 74 days, SD 56 days), and there were slightly more intervention participants who managed their diabetes with diet (230/368, 62.5%) than control group participants (207/355, 58.3%).

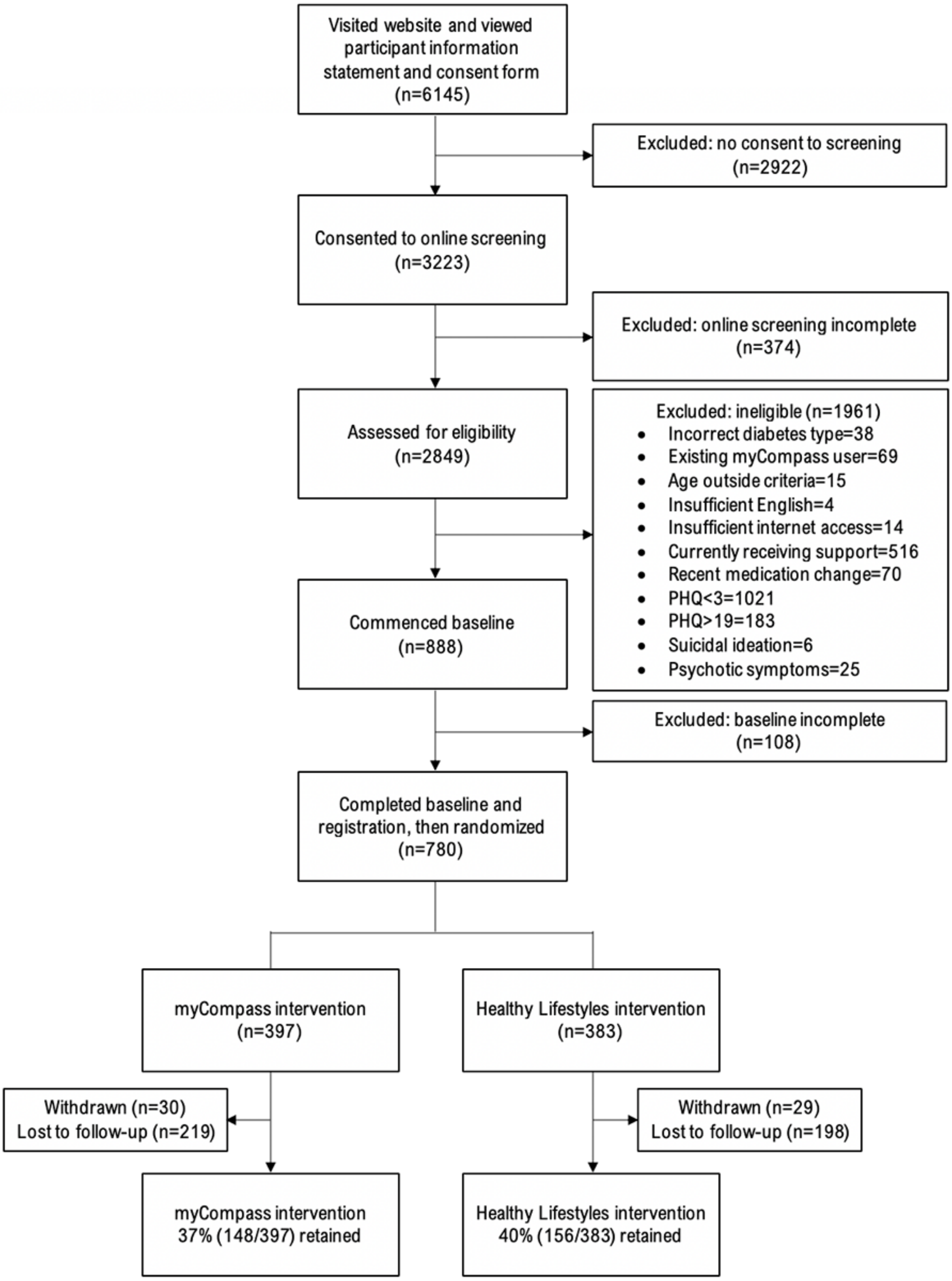
Table 2. Baseline means (SDs) for myCompass and Healthy Lifestyles Groups.

Characteristics	Percentage, n (%); (N=723)	myCompass (n=368), mean (SD)	Healthy lifestyles (n=355), mean (SD)	*P* value
Demographics				N/A[a]
Age (years)	N/A	57.7 (10.6)	57.7 (10.0)	
Female	465 (64.3)	229 (62)	236 (66)	
Married	387 (53.5)	204 (55)	183 (52)	
Employed	351 (48.5)	173 (47)	178 (50)	
Education level				N/A
Secondary school or lower	220 (30.4)	112 (30)	108 (30)	
Trade certificate or diploma	270 (37.3)	133 (36)	137 (39)	
University undergraduate or more	233 (32.2)	123 (33)	110 (31)	
Mental health				N/A
Lifetime history				
Sought professional support for mental health	571 (78.9)	296 (80)	275 (77)	
Received mental health diagnosis	300 (41.5)	155 (42)	145 (41)	
Diagnosed with depressive symptoms or major depressive disorder	279 (38.6)	143 (38)	136 (38)	
Past 6 weeks				N/A
Sought professional support for mental health	113 (15.6)	65 (18)	48 (14)	
Current				
Taking antidepressant medication	241 (33.3)	125 (34)	116 (33)	N/A
Months using antidepressant medication	N/A	97.70 (94.72)	73.67 (56.21)[b]	.04
Diabetes				
Age at diagnosis (years)	N/A	46.6 (11.1)	47.2 (10.9)	N/A
Diabetes treatment				
Healthy eating	437 (60.4)	230 (63)	207 (58)[b]	.04
Physical activity	323 (44.7)	176 (48)	147 (41)	N/A
Oral medication	583 (80.6)	295 (80)	288 (81)	N/A
Insulin	216 (29.9)	113 (31)	103 (29)	N/A
Exenatide	32 (4.4)	21 (6)	11 (3)	N/A
Past 6 weeks				N/A
Visited general practitioner for diabetes	419 (57.9)	218 (59)	201 (57)	
Frequency of general practitioner visit	N/A	1.31 (.78)	1.37 (.71)	
Hospitalized for diabetes	24 (3.3)	13 (4)	11 (3)	
Frequency of hospitalization for diabetes	N/A	1.46 (1.5)	1.36 (.9)	

[a] N/A: not applicable.

[b] Means differ significantly at *P*<.05

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) participant flow diagram through the SpringboarD trial. PHQ: Patient Health Questionnaire-9.



Intention-to-Treat Analyses (MMRM)

The estimated marginal means for all study outcomes are presented in Table 3, and the fixed effects from the MMRM analyses are given in Table 4. WSAS showed a significant improvement between pre- and postintervention, which was maintained at both 6- and 12-month follow-up. The effects of time differed significantly across the groups, with the Healthy Lifestyles group unexpectedly showing a greater improvement in work and social functioning across the trial than the myCompass group.

No other between-group differences were observed. All participants showed small but significant improvements on PHQ-9, DDS, and GAD-7 between baseline and postintervention, and these improvements were maintained across both follow-up periods. Similarly, all participants maintained significant increases in the healthy eating and blood glucose monitoring domains of the SMP-T2D across the trial, irrespective of group. HbA_{1c} decreased significantly between baseline and 6 months for all participants, but the reduction from baseline was no longer significant at 12 months.

Table 3. Estimated marginal means and SEs on key outcome variables for the myCompass and Healthy Lifestyle groups.

Variables and groups	Baseline, MC[a] (n=368), HL[b] (n=355), mean (SE)	3 months, MC (n=232), HL (n=241), mean (SE)	6 months, MC (n=216), HL (n=221), mean (SE)	12 months, MC (n=148), HL (n=156), mean (SE)
WSAS[c]				
HL	12.157 (0.456)	10.827 (0.523)	10.972 (0.538)	11.214 (0.542)
MC	13.440 (0.453)	12.055 (0.530)	11.926 (0.541)	11.642 (0.558)
PHQ-9[d]				
HL	10.802 (0.260)	8.411 (0.302)	8.379 (0.311)	8.017 (0.314)
MC	11.199 (0.258)	8.814 (0.307)	8.269 (0.314)	8.426 (0.325)
GAD-7[e]				
HL	7.236 (0.230)	6.231 (0.267)	6.322 (0.276)	5.929 (0.306)
MC	7.490 (0.228)	6.755 (0.271)	6.278 (0.277)	5.949 (0.317)
DDS[f]				
HL	2.536 (0.048)	2.237 (0.054)	2.205 (0.055)	2.203 (0.055)
MC	2.498 (0.047)	2.275 (0.054)	2.146 (0.055)	2.199 (0.057)
SMP-HE[g]				
HL	48.595 (1.522)	54.411 (1.742)	52.245 (1.788)	55.550 (1.804)
MC	50.909 (1.510)	53.206 (1.761)	55.053 (1.800)	52.951 (1.854)
SMP-BG[h]				
HL	45.705 (2.075)	53.760 (2.386)	51.664 (2.451)	50.991 (2.478)
MC	50.083 (2.058)	49.361 (2.425)	51.761 (2.473)	49.643 (2.545)
HbA_{1c}[i]				
HL	7.465 (0.089)	N/A[j]	7.255 (0.100)	7.256 (0.106)
MC	7.527 (0.088)	N/A	7.400 (0.103)	7.534 (0.109)

[a] MC: myCompass.

[b] HL: Healthy Lifestyles.

[c] WSAS: Work and Social Adjustment Scale.

[d] PHQ-9: Patient Health Questionnaire-9.

[e] GAD-7: Generalized Anxiety Disorder Scale.

[f] DDS: Diabetes Distress Scale.

[g] SMP-HE: Self-Management Profile for Type 2 Diabetes-Healthy Eating.

[h] SMP-BG: Self-Management Profile for Type 2 Diabetes-Blood Glucose Monitoring.

[i] HbA_{1c}: glycosylated hemoglobin (mmoL/L).

[j] N/A: not applicable.

Table 4. Mixed model repeated measures fixed effects for time, group, and time×group on primary and secondary outcome variables.

Variables and effects	β [a]	SE	95% CI	*df*	*t* test	*P* value
WSAS[b]						
Group	1.163	0.571	0.045 to 2.282	705.855	2.039	.04
T3[c]	–1.408	0.360	–2.113 to –0.703	1452.986	–3.913	<.001
T6[d]	–1.376	0.369	–2.100 to –0.653	1458.523	–3.729	<.001
T12[e]	–1.380	0.377	–2.118 to –0.642	1459.629	–3.663	<.001
PHQ-9[f]						
Group	.322	0.315	–0.295 to 0.939	716.640	1.022	.31
T3	–2.430	0.221	–2.863 to –1.998	1491.246	–11.009	<.001
T6	–2.703	0.226	–3.147 to –2.260	1498.607	–11.941	<.001
T12	–2.811	0.231	–3.265 to –2.358	1500.770	–12.157	<.001
GAD-7[g]						
Group	.223	0.281	–0.328 to 0.773	683.228	0.793	.43
T3	–0.926	0.195	–1.309 to –0.544	1365.156	–4.751	<.001
T6	–1.105	0.200	–1.498 to –0.712	1374.206	–5.513	<.001
T12	–1.472	0.224	–1.912 to –1.033	1386.943	–6.566	<.001
DDS[h]						
Group	–0.029	0.065	–0.157 to 0.099	692.095	–0.442	.66
T3	–0.272	0.034	–0.339 to –0.205	1391.445	–7.953	<.001
T6	–0.349	0.035	–0.418 to –0.280	1394.559	–9.935	<.001
T12	–0.325	0.036	–0.396 to –0.255	1394.854	–9.046	<.001
SMP-HE[i]						
Group	.944	1.913	–2.806 to 4.694	683.100	0.493	.62
T3	4.357	1.182	2.040 to 6.673	1419.606	3.686	<.001
T6	4.110	1.214	1.731 to 6.489	1424.907	3.386	<.001
T12	4.811	1.239	2.383 to 7.240	1425.917	3.883	<.001
SMP-BG[j]						
Group	.988	2.549	–4.009 to 5.984	692.279	0.387	.70
T3	3.998	1.651	0.762 to 7.234	1435.328	2.422	.02
T6	4.043	1.693	0.726 to 7.361	1441.001	2.389	.02
T12	2.683	1.728	–0.703 to 6.070	1442.227	1.553	.12
HbA_{1c}[k]						
Group	.139	0.118	–0.092 to 0.370	484.986	1.176	.24
T6	–0.169	0.068	–0.302 to –0.037	627.037	–2.507	.01
T12	–0.113	0.072	–0.255 to 0.029	638.526	–1.561	.12

[a] β is the unstandardized regression coefficient for the effect holding constant age, sex, years since diabetes diagnosis, use of diabetes medication (yes/no), and use of psychiatric medication (yes/no).

[b] WSAS: Work and Social Adjustment Scale.

[c] T3: 3-month follow-up.

[d] T6: 6-month follow-up.

[e] T12: 12-month follow-up.

[f] PHQ-9: Patient Health Questionnaire-9.

[g] GAD-7: Generalized Anxiety Disorder Scale.

[h] DDS: Diabetes Distress Scale.

[i] SMP-HE: Self-Management Profile for Type 2 Diabetes-Healthy Eating.

[j] SMP-BG: Self-Management Profile for Type 2 Diabetes-Blood Glucose Monitoring.

[k] HbA_{1c}: glycosylated hemoglobin (mmoL/L).

Change-Over-Time Analyses

Small but significant improvements in the functioning and well-being of the participants were observed over the course of the study. To explore the contribution of psychosocial factors to these improvements, we reran the MMRM analyses without the grouping factor (thereby estimating change over time for the full cohort) for all variables that did not differ by group (ie, all secondary outcomes). Each model contained all psychosocial variables simultaneously, so each effect was estimated while holding all other covariates constant. The unique effects of psychosocial variables on improvements over time are presented in Table 5.

Table 5. Mixed model repeated measure fixed effects of covariates on change over time in secondary outcome variables.

Covariate	PHQ-9[a]		GAD-7[b]		DDS[c]		SMP-HE[d]		SMP-BG[e]		HbA_{1c}[f]	
	β[g]	SE	β	SE	β	SE	β	SE	β	SE	β	SE
Age	–0.089[h]	0.016	–0.101[h]	0.014	–0.035[h]	0.003	0.579[h]	0.098	0.487[h]	0.132	–0.044[h]	0.006
Sex	0.381[h]	0.326	0.063	0.290	0.161	0.064	–3.421[i]	1.989	–6.145[i]	2.683	–0.026	0.119
Psychiatric meds	1.177	0.305	0.326	0.272	0.011	0.060	–1.584	1.864	–0.492	2.514	0.053	0.111
T2DM[j] meds	–0.178	0.388	0.017	0.346	0.007	0.076	–1.169	2.368	–6.488[i]	3.198	0.498[h]	0.142
Years since diagnosis	0.042[i]	0.020	0.017	0.018	0.012[h]	0.004	–0.136	0.123	–0.220	0.166	0.033[h]	0.007

[a] PHQ-9: Patient Health Questionnaire-9.

[b] GAD-7: Generalized Anxiety Disorder Scale.

[c] DDS: Diabetes Distress Scale.

[d] SMP-HE: Self-Management Profile for Type 2 Diabetes-Healthy Eating.

[e] SMP-BG: Self-Management Profile for Type 2 Diabetes-Blood Glucose Monitoring.

[f] HbA_{1c}: glycosylated hemoglobin (mmoL/L).

[g] β is the unstandardized regression coefficient for the effect.

[h] $P<.01$

[i] $P<.05$

[j] T2DM: type 2 diabetes.

Several psychosocial factors have demonstrated significant contributions to changes in the functioning and well-being over time. Age was the most consistent contributor, with each additional year of age associated with lower depression, anxiety, and diabetes distress; better dietary control and blood glucose level (BGL) monitoring; and better glycemic management. Being female was associated with higher depression over time and less time spent eating healthy food or BGL monitoring. Being on medication for T2DM was associated with less self-reported BGL monitoring but better glycemic management. Each additional year of holding a diabetes diagnosis was associated with higher depression and diabetes distress and poorer glycemic management.

Study Attrition

Of the total trial participants, 37.3% (148/397) of the participants from the myCompass program and 40.3% (156/387) of the participants from the Healthy Lifestyles program provided at least one measurement at the final time point (labeled *full completers*). To explore factors that affected engagement with the trial, we conducted a multivariate analysis of variance comparing noncompleters (participants who did not provide any postintervention measures) with full completers at baseline. Differences in diabetes distress and glycemic management were identified, with noncompleters reporting significantly more severe diabetes distress ($F_{1,783}$=6.784; P=.009; d=0.30) and higher HbA_{1c} ($F_{1,783}$=4.368; P=.04; d=0.24). Although these effects were small, it appears that diabetes distress and glycemic control were the main differentiating factors between those who completed the trial and those who did not.

Program Use and Feedback

myCompass participants logged in on an average of 6 times (SD 9.01; range 1-71), started an average of 0.71 modules (SD 1.18; range 0-8), completed an average of 0.29 modules (SD 0.87; range 0-7), and monitored their symptoms an average of 2 times (SD 5.79; range 0-53). Healthy Lifestyles participants logged in on an average of 4 times (SD 3.22; range 1-17), started an average of 2.61 modules (SD 2.78; range 0-8), and completed an average of 1.37 modules (SD 2.24; range 0-8). There were no differences between participants who logged into their assigned program and those who did not, apart from a slightly higher GAD-7 score reported by myCompass users who logged in (F=10.76, P=.001; d=0.39). No measure of program engagement correlated with any baseline measurement. With respect to program acceptability, approximately 55% of myCompass participants and 11% of Healthy Lifestyles participants reported that their assigned program was convenient and easy to use.

Discussion

Principal Findings

This trial examined the efficacy of an unguided eMH program (myCompass) for improving work and social functioning in people with T2DM and mild-to-moderate depressive symptoms, relative to an attention control program (Healthy Lifestyles). Contrary to our hypothesis, the Healthy Lifestyles group showed significantly greater improvement in work and social functioning than the myCompass group across the 12-month trial period. Irrespective of the intervention, all study participants reported significant improvements in their mental health and diabetes management by the end of the trial.

In our primary analysis that examined differences at our first (3-month) follow-up period [21], the Healthy Lifestyles group also showed a slightly greater improvement in medication adherence. Both these effects are surprising, given that the Healthy Lifestyles program has been a reliably inert control in previous trials [28]. Taken together, these effects suggest that the Healthy Lifestyles control intervention comprised more active ingredients than we had intended or was presented in a way that was particularly engaging to our chronically ill sample. Perhaps reflecting on general health prompted small behavioral changes, which gradually improved the overall functioning. Alternatively, a broad health literacy program may have been less confronting than a mental health program, which may be experienced as stigmatizing and adding greater health burden. This was suggested in a small qualitative study we previously conducted, in which young people with type I diabetes mellitus and T2DM reported that a focus on mental health negatively influenced their decision to take up and engage in eMH services [39]. The findings of this study provide a useful starting point for future research that should include further qualitative investigations of how and why people with diabetes use online mental health support.

All participants reported significant reductions in depression, anxiety, and diabetes distress throughout the trial, regardless of group. Participants also reported increases in the time spent eating healthily and monitoring blood glucose, along with small improvements in glycemic management, as measured by HbA_{1c}. These improvements over time did not meet the criteria for clinical significance but were robust across the sample and warrant consideration. It is an inescapable part of undertaking RCTs that require self-reported baseline measures that participants are assisted in reflecting on their own diabetes management, which can lead to improvements in diabetes health [38]. Being part of SpringboarD may have also created small increases in participants' health literacy across the trial, leading to commensurate changes in diabetes management. As mentioned earlier, the effects were generally small and require replication before any firm conclusions are drawn.

The relationship between self-reported diabetes management behavior and glycemic management in our sample was interesting. Both groups reported improvements in healthy eating and blood glucose monitoring throughout the trial; however, the total time spent on each remained at around 50% of the prior week. Nonetheless, all participants maintained HbA_{1c} levels within the recommended range of 7%-8% [40] throughout the trial. This finding was unexpected and suggested that some self-management behaviors may be effective even when not applied consistently throughout the week. Future research to establish the levels of self-management that are both practicable and effective may be fruitful, given the importance of lifestyle management in diabetes care [40]. Of course, such research should also consider other relevant factors not included in our analyses, such as diabetes medication effectiveness and adherence.

This study extended from our primary analyses [21] by examining the contribution of psychosocial factors to changes in the functioning and well-being over 12 months in adults with T2DM and mild depressive symptoms. In line with recent Australian data [41], age was a robust contributing factor, with being older associated with lower anxiety, depression, and stress; better self-management; and lower HbA_{1c}. Being female was associated with higher depression scores and poorer self-management. As may be expected, being on diabetes medication was associated with poorer BGL monitoring and higher HbA_{1c} likely related to longer duration and/or greater severity of disease.

Interestingly, having held a T2DM diagnosis for longer was associated with increased depression, diabetes distress, and HbA_{1c} levels, which may seem counterintuitive given that increased age was negatively associated with these factors. This also conflicts with previous data [42], indicating that both age and years since diagnosis predict poor glycemic management. Our findings suggest that although older people experience better functioning and well-being possibly related to health behaviors, living with T2DM longer is detrimental irrespective of age, possibly due to the worsening of cerebrovascular comorbidities over time [2]. Although recent data suggest that diabetes management may improve with age [41], the time postdiagnosis appears to be a separate and potentially detrimental factor in diabetes health. It may, therefore, be useful to consider these separately, both in future research and when assessing risk at the patient level.

Our lack of a treatment effect was surprising, given that myCompass showed promise as a treatment for depressive symptoms in a pilot trial of adults with diabetes [20], and myCompass has demonstrated efficacy in reducing depressive symptoms in the general community [19]. As discussed in our primary outcomes paper [21], this may be explained by methodological or population differences between previous work and this trial. Engagement in the myCompass program was lower than that observed in an earlier trial [19], giving rise to the possibility that people with diabetes may require or prefer more tailored interventions that directly address the challenges of diabetes management [41]. The lack of a treatment effect also likely reflects the impact of systematic attrition that resulted in near-floor mean baseline scores. Attrition is a well-known phenomenon in eMH research [43], and our data provide further impetus for an ongoing discussion of methodology in this area.

Strengths and Limitations

Our challenges with recruitment and attrition, although not ideal, were ultimately informative. Participants with the highest symptoms of distress and impairment tended to leave the study. As a result, near-normal scores on baseline variables weakened tests of efficacy, and our results are largely indicative of people with T2DM and mild distress or impairment. Nonetheless, the inclusion of multiple follow-up assessments in this analyses and large sample that our study comprised afforded us the opportunity to extend the trial findings by analyzing recruitment strategies [23] and examining study attrition in both short and long terms.

At our first follow-up point, people who had left the study were characterized by more severe depressive and anxiety symptoms, greater diabetes-related distress, and poorer medication adherence [21]. However, by the final follow-up point, study-leavers were differentiated from study-completers by increased diabetes distress and poorer glycemic management. This suggests that in diabetes-related trials, short-term retention may be impacted by psychiatric factors (such as depression or anxiety), whereas diabetes-specific factors (such as emotional adjustment to diabetes) may impact both short- and long-term retention. This may reflect overall fatigue in chronic disease management.

Our insights could be valuable for future research. Retention strategies for future studies may need to vary across study phases, and the impact of this apparently biphasic pattern of attrition could be taken into account when analyzing results. In addition, future studies could continue this contribution to methodological improvements by further investigating factors influencing recruitment and trial engagement. For example, the association between age and improved functioning suggests that future recruitment strategies should focus on younger participants to ensure interventions are trialed with those who most need support and, therefore, are most likely to derive measurable benefits.

Conclusions

The SpringboarD trial aimed to determine if a public health eMH program, myCompass, could improve the work and social functioning in adults with T2DM and mild-to-moderate depressive symptoms. The trial also sought to examine the impact of myCompass on a range of physical and mental health outcomes. There was a small, unexpected benefit to our control group in terms of work and social functioning, which suggests the need for future research to examine the value of generic health literacy tools. Neither control nor intervention programs were found to yield specific mental health benefits, but systematic attrition likely hampered true tests of efficacy by yielding only a mildly symptomatic sample.

Nonetheless, the trial itself revealed valuable insights into studying mental health in the context of T2DM. Early-stage attrition seems to be affected by mental health, whereas late-stage attrition seems more impacted by diabetes health. Increasing age appears to be associated with a gradual lift in both mental and physical health in T2DM; thus, research may show benefits from targeting younger cohorts. As mental health continues to be a significant contributor to morbidity in diabetes, increasingly refined approaches are required to meet the sizable demand for mental health support in people with T2DM.

Acknowledgments

The authors gratefully acknowledge the dedication and input of all the study participants. The authors also acknowledge the support of the Australian Government Department of Health in funding the development and dissemination of the myCompass program and the support of the National Health and Medical Research Council (NHMRC) in funding this research trial (NHMRC Grant 1083116).

Conflicts of Interest

JP was involved in the development of myCompass.

References

1. Nanayakkara N, Pease A, Ranasinha S, Wischer N, Andrikopoulos S, Speight J, et al. Depression and diabetes distress in adults with type 2 diabetes: results from the Australian national diabetes audit (ANDA) 2016. Sci Rep 2018 May 18;8(1):7846 [FREE Full text] [doi: 10.1038/s41598-018-26138-5] [Medline: 29777153]
2. Biessels GJ, Staekenborg S, Brunner E, Brayne C, Scheltens P. Risk of dementia in diabetes mellitus: a systematic review. Lancet Neurol 2006 Jan;5(1):64-74. [doi: 10.1016/S1474-4422(05)70284-2] [Medline: 16361024]
3. Rubin RR, Peyrot M. Psychological issues and treatments for people with diabetes. J Clin Psychol 2001 Apr;57(4):457-478. [doi: 10.1002/jclp.1041] [Medline: 11255202]
4. Semenkovich K, Brown ME, Svrakic DM, Lustman PJ. Depression in type 2 diabetes mellitus: prevalence, impact, and treatment. Drugs 2015 Apr;75(6):577-587. [doi: 10.1007/s40265-015-0347-4] [Medline: 25851098]
5. Roy T, Lloyd CE. Epidemiology of depression and diabetes: a systematic review. J Affect Disord 2012 Oct;142(Suppl):S8-21. [doi: 10.1016/S0165-0327(12)70004-6] [Medline: 23062861]
6. Renn BN, Feliciano L, Segal DL. The bidirectional relationship of depression and diabetes: a systematic review. Clin Psychol Rev 2011 Dec;31(8):1239-1246. [doi: 10.1016/j.cpr.2011.08.001] [Medline: 21963669]

7. Li C, Xu D, Hu M, Tan Y, Zhang P, Li G, et al. A systematic review and meta-analysis of randomized controlled trials of cognitive behavior therapy for patients with diabetes and depression. J Psychosom Res 2017 Apr;95:44-54. [doi: 10.1016/j.jpsychores.2017.02.006] [Medline: 28314548]
8. Pols AD, Schipper K, Overkamp D, van Marwijk HW, van Tulder MW, Adriaanse MC. Patients' and practice nurses' perceptions of depression in patients with type 2 diabetes and/or coronary heart disease screened for subthreshold depression. BMC Fam Pract 2018 Dec 23;19(1):202 [FREE Full text] [doi: 10.1186/s12875-018-0870-y] [Medline: 30579329]
9. Schmitz N, Gariépy G, Smith KJ, Clyde M, Malla A, Boyer R, et al. Recurrent subthreshold depression in type 2 diabetes: an important risk factor for poor health outcomes. Diabetes Care 2014 Apr;37(4):970-978. [doi: 10.2337/dc13-1832] [Medline: 24198303]
10. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. Use of mental health services in Europe: results from the European study of the epidemiology of mental disorders (ESEMeD) project. Acta Psychiatr Scand Suppl 2004(420):47-54. [doi: 10.1111/j.1600-0047.2004.00330.x] [Medline: 15128387]
11. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. Health Aff (Millwood) 2009;28(3):w490-w501. [doi: 10.1377/hlthaff.28.3.w490] [Medline: 19366722]
12. Andrews G, Cuijpers P, Craske M, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
13. Matcham F, Rayner L, Hutton J, Monk A, Steel C, Hotopf M. Self-help interventions for symptoms of depression, anxiety and psychological distress in patients with physical illnesses: a systematic review and meta-analysis. Clin Psychol Rev 2014 Mar;34(2):141-157 [FREE Full text] [doi: 10.1016/j.cpr.2014.01.005] [Medline: 24508685]
14. van Bastelaar KM, Pouwer F, Cuijpers P, Riper H, Snoek FJ. Web-based depression treatment for type 1 and type 2 diabetic patients: a randomized, controlled trial. Diabetes Care 2011 Feb;34(2):320-325 [FREE Full text] [doi: 10.2337/dc10-1248] [Medline: 21216855]
15. Newby J, Robins L, Wilhelm K, Smith J, Fletcher T, Gillis I, et al. Web-based cognitive behavior therapy for depression in people with diabetes mellitus: a randomized controlled trial. J Med Internet Res 2017 May 15;19(5):e157 [FREE Full text] [doi: 10.2196/jmir.7274] [Medline: 28506956]
16. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
17. Fisher L, Mullan JT, Arean P, Glasgow RE, Hessler D, Masharani U. Diabetes distress but not clinical depression or depressive symptoms is associated with glycemic control in both cross-sectional and longitudinal analyses. Diabetes Care 2010 Jan;33(1):23-28 [FREE Full text] [doi: 10.2337/dc09-1238] [Medline: 19837786]
18. Fisher L, Gonzalez JS, Polonsky WH. The confusing tale of depression and distress in patients with diabetes: a call for greater clarity and precision. Diabet Med 2014 Jul;31(7):764-772 [FREE Full text] [doi: 10.1111/dme.12428] [Medline: 24606397]
19. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
20. Clarke J, Proudfoot J, Ma H. Mobile phone and web-based cognitive behavior therapy for depressive symptoms and mental health comorbidities in people living with diabetes: results of a feasibility study. JMIR Ment Health 2016 May 31;3(2):e23 [FREE Full text] [doi: 10.2196/mental.5131] [Medline: 27245948]
21. Clarke J, Sanatkar S, Baldwin PA, Fletcher S, Gunn J, Wilhelm K, et al. A web-based cognitive behavior therapy intervention to improve social and occupational functioning in adults with type 2 diabetes (the springboard trial): randomized controlled trial. J Med Internet Res 2019 May 21;21(5):e12246 [FREE Full text] [doi: 10.2196/12246] [Medline: 31115345]
22. Proudfoot J, Clarke J, Gunn J, Fletcher S, Sanatkar S, Wilhelm K, et al. A web-based public health intervention to reduce functional impairment and depressive symptoms in adults with type 2 diabetes (the Springboard trial): randomized controlled trial protocol. JMIR Res Protoc 2017 Aug 3;6(8):e145 [FREE Full text] [doi: 10.2196/resprot.7348] [Medline: 28778848]
23. Fletcher S, Clarke J, Sanatkar S, Baldwin P, Gunn J, Zwar N, et al. Recruiting to a randomized controlled trial of a web-based program for people with type 2 diabetes and depression: lessons learned at the intersection of e-mental health and primary care. J Med Internet Res 2019 May 24;21(5):e12793 [FREE Full text] [doi: 10.2196/12793] [Medline: 31127718]
24. 45Up Study Collaborators, Banks E, Redman S, Jorm L, Armstrong B, Bauman A, et al. Cohort profile: the 45 and up study. Int J Epidemiol 2008 Oct;37(5):941-947 [FREE Full text] [doi: 10.1093/ije/dym184] [Medline: 17881411]
25. Kroenke K, Spitzer RL, Williams JB. The patient health questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
26. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

27. Degenhardt L, Hall W, Korten A, Jablensky A. Use of Brief Screening Instrument for Psychosis: Results of a ROC Analysis. National Drug and Alcohol Research Centre (NDARC). 2005. URL: https://ndarc.med.unsw.edu.au/resource/use-brief-screening-instrument-psychosis-results-roc-analysis [accessed 2020-08-11]
28. Christensen H, Griffiths KM, Mackinnon AJ, Kalia K, Batterham PJ, Kenardy J, et al. Protocol for a randomised controlled trial investigating the effectiveness of an online e health application for the prevention of generalised anxiety disorder. BMC Psychiatry 2010 Mar 21;10:25 [FREE Full text] [doi: 10.1186/1471-244X-10-25] [Medline: 20302678]
29. Mundt JC, Marks IM, Shear MK, Greist JH. The work and social adjustment scale: a simple measure of impairment in functioning. Br J Psychiatry 2002 May;180:461-464. [doi: 10.1192/bjp.180.5.461] [Medline: 11983645]
30. Pedersen G, Kvarstein EH, Wilberg T. The work and social adjustment scale: psychometric properties and validity among males and females, and outpatients with and without personality disorders. Personal Ment Health 2017 Nov;11(4):215-228. [doi: 10.1002/pmh.1382] [Medline: 28681505]
31. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
32. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
33. van Steenbergen-Weijenburg KM, de Vroege L, Ploeger RR, Brals JW, Vloedbeld MG, Veneman TF, et al. Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. BMC Health Serv Res 2010 Aug 12;10:235 [FREE Full text] [doi: 10.1186/1472-6963-10-235] [Medline: 20704720]
34. Polonsky WH, Fisher L, Earles J, Dudl RJ, Lees J, Mullan J, et al. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. Diabetes Care 2005 Mar;28(3):626-631. [doi: 10.2337/diacare.28.3.626] [Medline: 15735199]
35. Peyrot M, Bushnell DM, Best JH, Martin ML, Cameron A, Patrick DL. Development and validation of the self-management profile for type 2 diabetes (SMP-T2D). Health Qual Life Outcomes 2012 Oct 5;10:125 [FREE Full text] [doi: 10.1186/1477-7525-10-125] [Medline: 23039868]
36. Üstün T, Kostanjsek N, Chatterji S, Rehm J. Measuring Health and Disability: Manual for WHO Disability Assessment Schedule WHODAS 2.0. Geneva, Switzerland: World Health Organization; 2019.
37. R: A Language and Environment for Statistical Computing. The R Project for Statistical Computing. URL: http://www.R-project.org/ [accessed 2019-07-06]
38. Nazar CM, Bojerenu MM, Safdar M, Marwat J. Effectiveness of diabetes education and awareness of diabetes mellitus in combating diabetes in the United Kingdom; a literature review. J Nephropharmacol 2016;5(2):110-115 [FREE Full text] [Medline: 28197516]
39. Clarke J, Proudfoot J, Vatiliotis V, Verge C, Holmes-Walker DJ, Campbell L, et al. Attitudes towards mental health, mental health research and digital interventions by young adults with type 1 diabetes: a qualitative analysis. Health Expect 2018 Jun;21(3):668-677 [FREE Full text] [doi: 10.1111/hex.12662] [Medline: 29319923]
40. American Diabetes Association. Standards of medical care in diabetes--2013. Diabetes Care 2013 Jan;36(Suppl 1):S11-S66 [FREE Full text] [doi: 10.2337/dc13-S011] [Medline: 23264422]
41. Nanayakkara N, Ranasinha S, Gadowski AM, Davis WA, Flack JR, Wischer N, et al. Age-related differences in glycaemic control, cardiovascular disease risk factors and treatment in patients with type 2 diabetes: a cross-sectional study from the Australian national diabetes audit. BMJ Open 2018 Aug 17;8(8):e020677 [FREE Full text] [doi: 10.1136/bmjopen-2017-020677] [Medline: 30121593]
42. Rady MY, Johnson DJ, Patel BM, Larson JS, Helmers RA. Influence of individual characteristics on outcome of glycemic control in intensive care unit patients with or without diabetes mellitus. Mayo Clin Proc 2005 Dec;80(12):1558-1567. [doi: 10.4065/80.12.1558] [Medline: 16342648]
43. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

Abbreviations

BGL: blood glucose level
CBT: cognitive behavioral therapy
DDS: Diabetes Distress Scale
eMH: electronic mental health
GAD-7: Generalized Anxiety Disorder Questionnaire-7
GP: general practitioner
HbA$_{1c}$: glycosylated hemoglobin
HREC: Human Research Ethics Committee
MMRM: mixed model repeated measure
NHMRC: National Health and Medical Research Council
PHQ-9: Patient Health Questionnaire-9 item

RCT: randomized controlled trial
SMP-T2D: Self-Management Profile for Type 2 Diabetes
T2DM: type 2 diabetes
WSAS: Work and Social Adjustment Scale

Edited by G Eysenbach; submitted 17.10.19; peer-reviewed by J Huber, L Schwab-Reese, H Jin; comments to author 28.11.19; revised version received 26.02.20; accepted 22.06.20; published 01.12.20.

Please cite as:
Baldwin PA, Sanatkar S, Clarke J, Fletcher S, Gunn J, Wilhelm K, Campbell L, Zwar N, Harris M, Lapsley H, Hadzi-Pavlovic D, Christensen H, Proudfoot J
A Web-Based Mental Health Intervention to Improve Social and Occupational Functioning in Adults With Type 2 Diabetes (The Springboard Trial): 12-Month Outcomes of a Randomized Controlled Trial
J Med Internet Res 2020;22(12):e16729
URL: https://www.jmir.org/2020/12/e16729
doi:10.2196/16729
PMID:33258790

Original Paper

A Mobile- and Web-Based Health Intervention Program for Diabetes and Prediabetes Self-Management (BetaMe/Melon): Process Evaluation Following a Randomized Controlled Trial

Virginia Signal[1], PhD; Melissa McLeod[1], PhD; James Stanley[2], PhD; Jeannine Stairmand[1], MPH; Nitin Sukumaran[3], DPH; Donna-Marie Thompson[1], BN; Kelly Henderson[1], BN; Cheryl Davies[4]; Jeremy Krebs[5], MD; Anthony Dowell[6], MBChB; Rebecca Grainger[5], PhD; Diana Sarfati[1], PhD

[1]Department of Public Health, University of Otago Wellington, Wellington, New Zealand
[2]Biostatistical Group, Dean's Department, University of Otago Wellington, Wellington, New Zealand
[3]School of Medicine, University of Otago, Dunedin, New Zealand
[4]Kokiri Marae, Wellington, New Zealand
[5]Department of Medicine, University of Otago Wellington, Wellington, New Zealand
[6]Department of Primary Health Care and General Practice, University of Otago Wellington, Wellington, New Zealand

Corresponding Author:
Virginia Signal, PhD
Department of Public Health
University of Otago Wellington
23a Mein Street
Newtown
Wellington,
New Zealand
Phone: 64 212631385
Email: virginia.signal@otago.ac.nz

Abstract

Background: Technology-assisted self-management programs are increasingly recommended to patients with long-term conditions such as diabetes. However, there are a number of personal and external factors that affect patients' abilities to engage with and effectively utilize such programs. A randomized controlled trial of a multi-modal online program for diabetes self-management (BetaMe/Melon) was conducted in a primary care setting, and a process evaluation was completed at the end of the study period.

Objective: This process evaluation aimed to examine the utilization patterns of BetaMe/Melon, identify which components participants found most (and least) useful, and identify areas of future improvement.

Methods: Process evaluation data were collected for intervention arm participants from 3 sources: (1) the mobile/web platform (to identify key usage patterns over the 16-week core program), (2) an online questionnaire completed during the final study assessment, and (3) interviews conducted with a subset of participants following the study period. Participants were classified as "actively engaged" if any usage data was recorded for the participant (in any week), and patterns were reported by age, gender, ethnicity, and diabetes/prediabetes status. The online questionnaire asked participants about the usefulness of the program and whether they would recommend BetaMe/Melon to others according to a 5-point Likert Scale. Of 23 invited participants, 18 participated in a digitally recorded, semistructured telephone interview. Interview data were thematically analyzed.

Results: Out of the 215 participants, 198 (92%) received an initial health coaching session, and 160 (74%) were actively engaged with the program at some point during the 16-week core program. Engagement varied by demographic, with women, younger participants, and ethnic majority populations having higher rates of engagement. Usage steadily declined from 50% at Week 0 to 23% at Week 15. Participants ranked component usefulness as education resources (63.7%), health coaches (59.2%), goal tracking (48.8%), and online peer support (42.1%). Although 53% agreed that the program was easy to use, 64% would recommend the program to others. Interview participants found BetaMe/Melon useful overall, with most identifying beneficial outcomes such as increased knowledge, behavioral changes, and weight loss. Barriers to engagement were program functionality, internet connectivity, incomplete delivery of all program components, and participant motivation. Participants suggested a range of improvements to the BetaMe/Melon program.

Conclusions: The program was generally well received by participants; active engagement was initially high, although it declined steadily. Maintaining participant engagement over time, individualizing programs, and addressing technical barriers are important to maximize potential health benefits from online diabetes self-management programs.

Trial Registration: Australian New Zealand Clinical Trial Registry ACTRN12617000549325; https://tinyurl.com/y622b27q

(J Med Internet Res 2020;22(12):e19150) doi:10.2196/19150

KEYWORDS

diabetes mellitus; prediabetes; self-management; eHealth; mobile apps; evaluation; diabetes; digital health; app; utilization; user perception; user

Introduction

Long-term conditions pose a great burden to patients and health services [1,2]. Furthermore, the economic burden on health systems is growing globally [3-6]. Self-management interventions are a potential way to address this burden, with evidence that such interventions can effectively improve users' physical and mental health [7]. Digital health interventions are increasingly available to support patient self-management [8-10]. However, there are personal and external factors that affect a person's ability to engage with and effectively utilize digital health interventions. These include age, motivation, personal values, lifestyle, digital literacy, and support from family and peers [11,12]. Factors external to an individual that influence engagement include the quality of the digital health intervention itself, internet access, level of support provided to enroll and participate, cost to the participant, clinical endorsement, and participant perceptions of data safety [11,12]. The degree to which programs have been designed using contextually relevant models and theory that target the desired behavior (eg, the Behaviour Change Wheel [13]) may be important to outcomes [14].

The BetaMe/Melon digital health intervention is a comprehensive mobile and web-based technology program for people with type 2 diabetes or prediabetes that uses principles of behavioral change theory to support and enhance users' self-management techniques [15]. The program was developed in partnership with primary care practitioners, Māori and Pacific health providers, and psychologists, and piloted with people with prediabetes [15]. The 12-month program has an initial 16-week active support phase comprising four intervention components: (1) health coaching (during which the first session with the health coach is the only compulsory component of the program), (2) goal setting and tracking, (3) online peer support in a secure forum, and (4) provision of evidence-based resources. The remaining 36 weeks use online peer support and goal tracking only.

Although multimodal digital health intervention programs such as BetaMe/Melon have been shown to be effective in supporting users' glycemic control [16-19], there are still many knowledge gaps regarding the use and efficacy of digital health interventions. For example, little is known about the mechanisms that might enhance individual engagement and thus have the greatest impact in terms of utilization and adherence [11,12,20,21]. There is also a lack of high-quality empirical evidence as to how individuals incorporate digital health intervention programs into their daily lives, which program components they find most helpful, and how they feel such programs could be improved [9,21].

Process evaluation is an essential part of assessing complex interventions [22,23] and can provide information about how an intervention might work in a given context, factors that may have impacted an intervention's outcomes, and modifiable factors that might improve an intervention's success [22]. The aims of our process evaluation were to assess how consistently and in what ways participants used the BetaMe/Melon mobile/web platform and to identify which program components participants found to be the most (and least) useful so as to identify future areas of improvement for this and other similar digital health intervention programs.

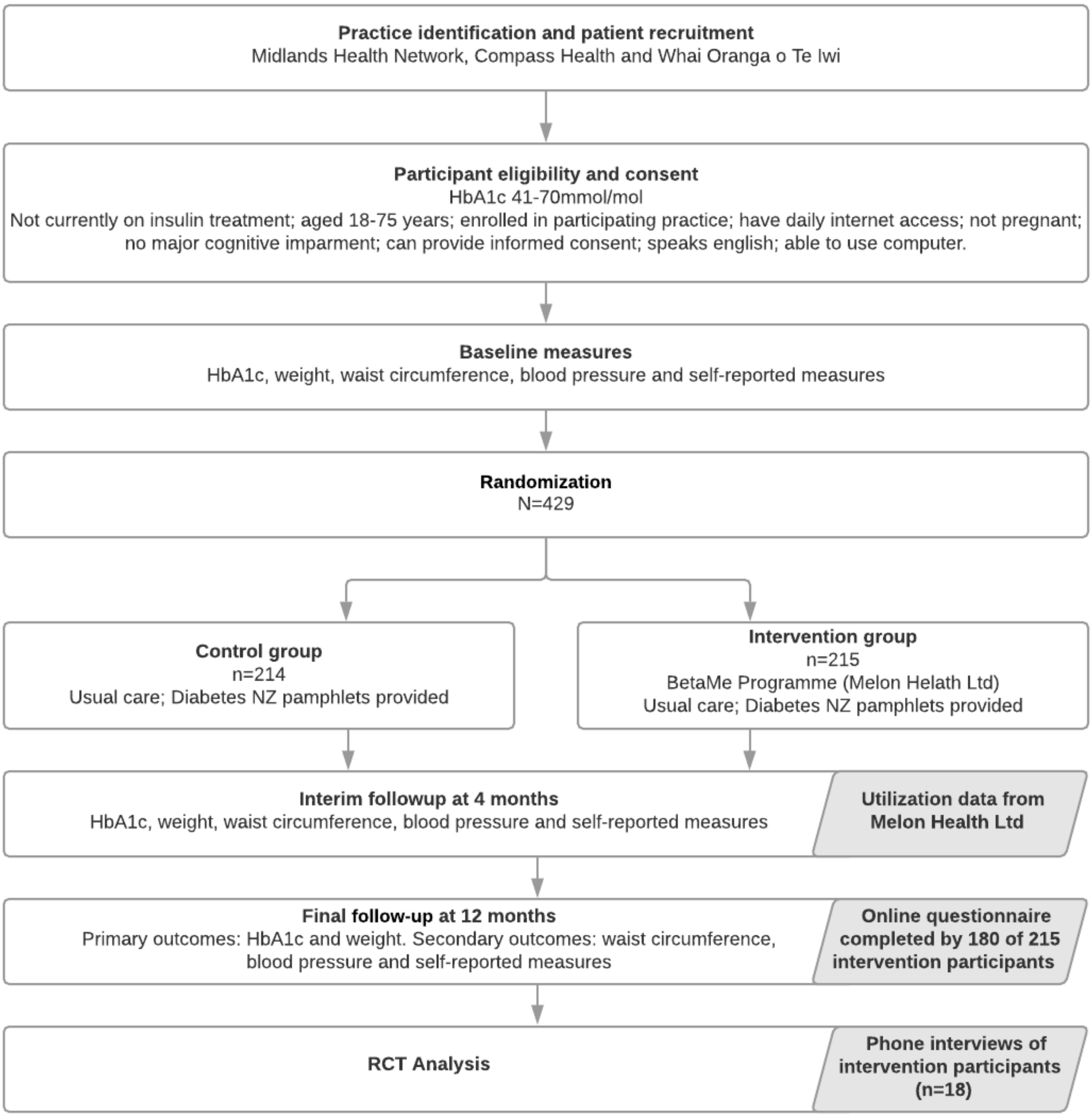
Methods

Design

We undertook a randomized controlled trial (RCT) of the BetaMe/Melon program with the design and methods described elsewhere [15]. Briefly, the RCT was carried out from June 2016 to June 2018 and compared the outcomes for people aged 18-75 years with hemoglobin A_{1c} (HbA_{1c}) of 41-70 mmol/mol enrolled in the BetaMe/Melon program or receiving usual primary health care. The RCT recruited 429 participants, with half (n=215) randomly allocated to the intervention arm (BetaMe/Melon) and half (n=214) allocated to the control arm (usual primary health care) of the study. The primary outcomes were mean changes in HbA_{1c} and weight from baseline to 12 months. A number of secondary outcomes were also measured [15].

Process evaluation data were collected for intervention arm participants from three sources: (1) the mobile/web platform (to identify key usage patterns during the 16-week active support phase), (2) an online questionnaire completed during the final study assessment, and (3) semistructured telephone interviews conducted with a subset of participants at the end of the study period. A study flow diagram highlighting when data were collected and analyses conducted is provided (Figure 1).

Figure 1. Study flow diagram. HbA_{1C}: hemoglobin A_{1C}; RCT: randomized controlled trial.



Ethical Approval

Ethical approval was given by the New Zealand Health and Disability Ethics Committee (approval reference: HDEC 17/CEN/49). The process evaluation was conducted by researchers from the University of Otago. All design, implementation, analytic, and dissemination aspects of the RCT and this process evaluation were the sole responsibility of the Otago research team. The research team had no financial relationship with the company (Melon Health Ltd) who developed the BetaMe/Melon program and delivered the program for the study under contract to the University.

Participants gave written consent to participate in the RCT, which included the use of information from the mobile/web platform to evaluate the BetaMe/Melon program. Additional verbal consent was gained from those participating in the telephone interviews at the end of the study period.

Utilization Patterns

Online usage data were provided by Melon Health to the University of Otago research team for analysis. The provided dataset covered the program’s first 16 weeks (active support phase) and included data on the first health coach session and all events where participants made an online diary entry or actively engaged with the peer support forum (posted, replied to, or “liked” messages). No data were available on passive interactions with the online program (eg, reading forum messages without commenting or liking) or use of educational resources.

For the utilization analysis, usage data were linked to demographic data collected during the baseline study assessment of the RCT. Participants were classified as "actively engaged" if any usage data was recorded for the participant (in any week), while participants with no usage data recorded were classified as "not engaged" during the active support phase. Proportions engaged are reported with 95% CIs, with differences between key demographic and clinical groups calculated as absolute differences with 95% CIs (differences not adjusted for other variables). Weekly engagement with the online program components was assessed for the active support phase (Week 0-16 with Week 0 starting at the time of enrollment) using descriptive statistics.

Utilization patterns are reported by demographic characteristics of age group (35-45, 45-55, 65-75 years), gender (male, female), self-identified ethnicity (using a prioritized order as follows: Māori, Pacific, or non-Māori/non-Pacific), and study group (diabetes/prediabetes). Diabetes/prediabetes status was defined using HbA_{1c} levels measured at the start of the study (diabetes range: 50-70 mmol/mol; prediabetes range: 41-49 mmol/mol). Utilization analyses were conducted and plotted in R3.5 (R Institute, Vienna, Austria).

Participant Perceptions

Online Survey

An online questionnaire was completed within REDCap (Research Electronic Data Capture, an electronic data capture tool hosted at the University of Otago [24,25]) by intervention arm participants during the RCT's final study assessment at 12 months after enrollment. Participants were asked about the usefulness of the full 12-month program and whether they would recommend the BetaMe/Melon program to others. Responses were collected using a 5-point Likert Scale (strongly disagree, disagree, neutral, agree, strongly agree). Data from the questionnaire were extracted from REDCap, analyzed in Microsoft Excel, and reported with descriptive statistics.

Participant Interviews

Interview participants were purposefully sampled to highlight Māori and Pacific peoples' views, with all Māori participants (n=9) and all Pacific participants (n=5) who agreed to be re-contacted invited to interview. The 18 interviews that were completed included all Māori and all Pacific participants who consented and a random sample of participants of other ethnicities stratified by primary care practice. An information sheet and consent form were emailed to all consenting participants prior to the interview.

Telephone interviews were conducted by a researcher with extensive experience in qualitative research (JSt). A semistructured interview schedule (see Multimedia Appendix 1) was used to elicit in-depth participant perceptions about the most and least useful components of the 12-month program, any barriers to incorporating BetaMe/Melon into daily life, and suggested changes to the program. All interviews were recorded digitally.

Participant interview data were analyzed thematically using a primarily inductive approach (led by JSt in close collaboration with VS). First, all interviews were listened to several times so as to familiarize analysists with the data. Second, a summary of findings for each participant was entered into an Excel spreadsheet by interview question. Third, interesting and important data features were coded and a coding framework developed. Fourth, initial themes were developed and reviewed against the dataset. Finally, findings and issues were discussed with other research team members and themes defined and named.

Results

Utilization Patterns

Of the 215 intervention participants, 92% (n=198) received an initial health coaching session. Active engagement (at any point in the 16-week monitored period) is shown in Table 1. Of the 215 participants, 160 (74%, 95% CI 68.0-80.1) were actively engaged with the online program components at some point during the 16-week active support phase; the remaining 55 participants (26%) did not actively engage with the online components at any time during the 16 weeks.

Patterns of any engagement in the 16-week period were broadly similar across subgroups. Although there was some variation between groups, much of this variation may be explained by relatively small sample sizes in subgroups rather than representing systematic tendencies for one type of participant to engage more than others (as seen with the CIs for proportions and their absolute differences in Table 1). The exception was engagement by gender, where women were more likely to actively engage at any time (82.4%, 89/108) than men (66.4%, 71/107). For other key comparisons, the differences were inconclusive.

Table 1. Patterns of any engagement (at any time in program) by key participant characteristics.

Participant characteristics	Total (N=215), n	Any engagement[a], n (%), 95% CI	Absolute difference[b] (95% CI)
Age (years)			
35-54	43	34 (79.1), 64.0-90.0	9.3 (–6.6-25.3)
55-64	76	53 (69.7), 58.1-79.8	Reference
65-74	96	73 (76.0), 66.3-84.2	6.3 (–7.1-19.7)
Ethnicity			
Māori	32	21 (65.6), 46.8-81.4	–10.1 (–27.7-7.5)
Pacific	6	5 (83.3), 35.9-99.6	7.6 (–22.9-38.1)
Non-Māori / non-Pacific	177	5 (75.7), 68.7-81.8	Reference
Gender			
Male	108	89 (82.4), 73.9-89.1	16.0 (4.6-27.5)
Female	107	71 (66.4), 56.6-75.2	Reference
Diabetes status			
Prediabetes	105	82 (78.1), 9.0-85.6	Reference
Diabetes	110	78 (70.9), 61.5-79.2	–7.2 (–18.8-4.4)

[a] Defined as at least one active engagement on the online portal, at any time.

[b] Absolute difference in proportion engaged relative to reference group

Overall, there was a steady decline in utilization rates (proportion of enrolled individuals actively engaged with the online program in each week) over the 16-week active support phase, from 50% (108/215) of participants engaging at Week 0 down to 22.3% (48/215) at Week 15. Figure 2 shows usage patterns over time by age group, gender, ethnicity, and diabetes status, with all groups showing higher engagement at Week 0 with a steady decline during active program delivery (Weeks 0-15). Participants in the youngest age group (35-54 years) had higher engagement than other age groups, but the trajectory for this age group quickly (by around Week 6) converged with engagement rates for other age groups. Female follow-up engagement was higher, and the percentage of non-Māori/non-Pacific participants engaged at any time point was higher compared to Māori participants. Engagement by participants identifying as Māori decreased the most rapidly to only 3% (1/32) engaging at the end of the 16-week period compared to 27% of non-Māori/non-Pacific participants (47/177). Engagement trajectories were similar for participants with initial HbA_{1c} in the diabetes and prediabetes ranges, following the overall pattern of results. Diary entries (eg, tracking progress toward health goals or changes in weight) were the most frequent form of engagement by all participants (Figure 3).

Figure 2. Participant active engagement with BetaMe/Melon portal over time, according to key participant characteristics (age group, gender, ethnicity*, and diabetes status group). *Results for Pacific participants are not presented due to small sample size (n=6).

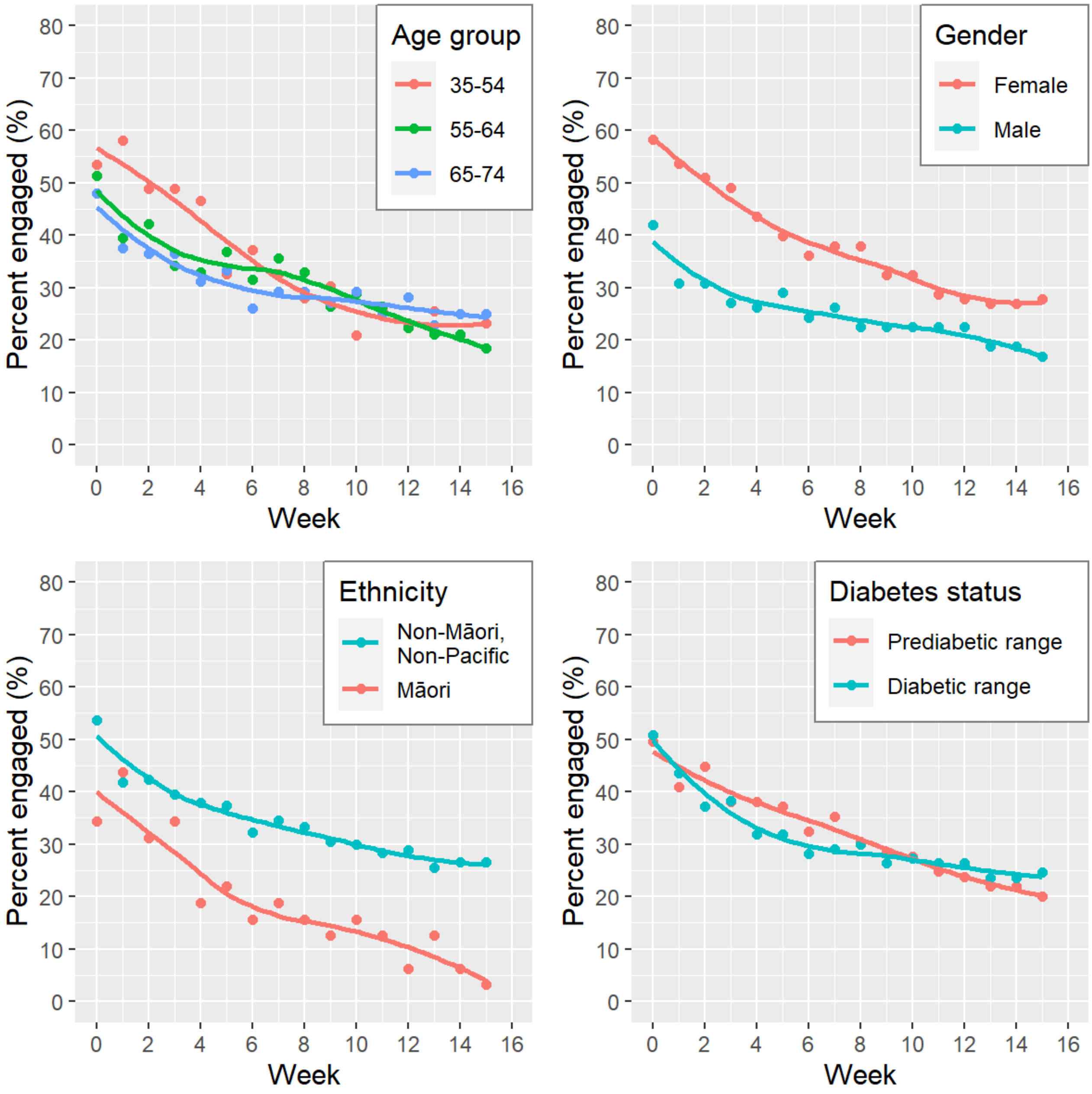
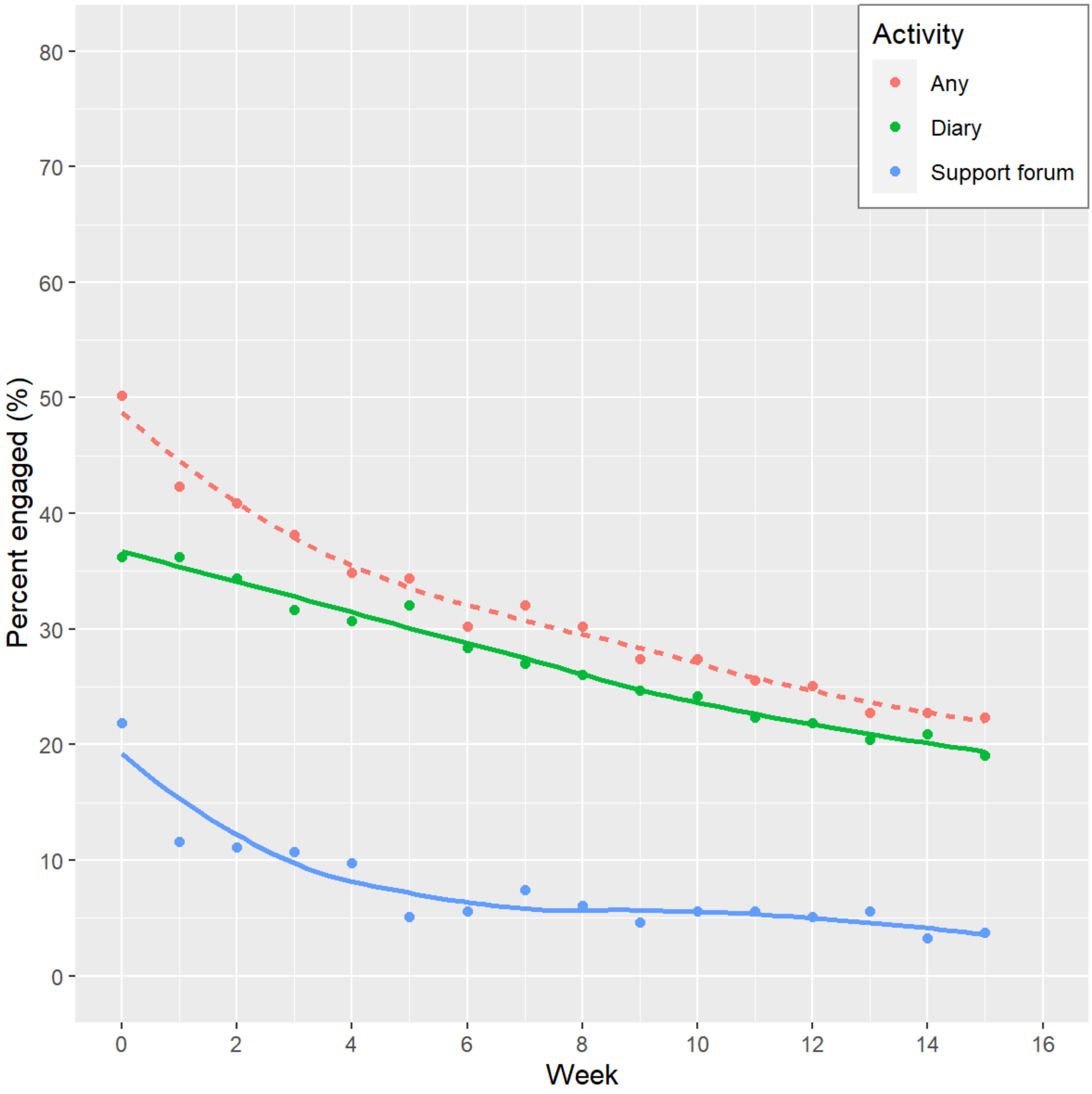


Figure 3. Trends in engagement by week according to type of activity (diary, online peer support forum). "Any" activity represents percentage with any active engagement in that week.

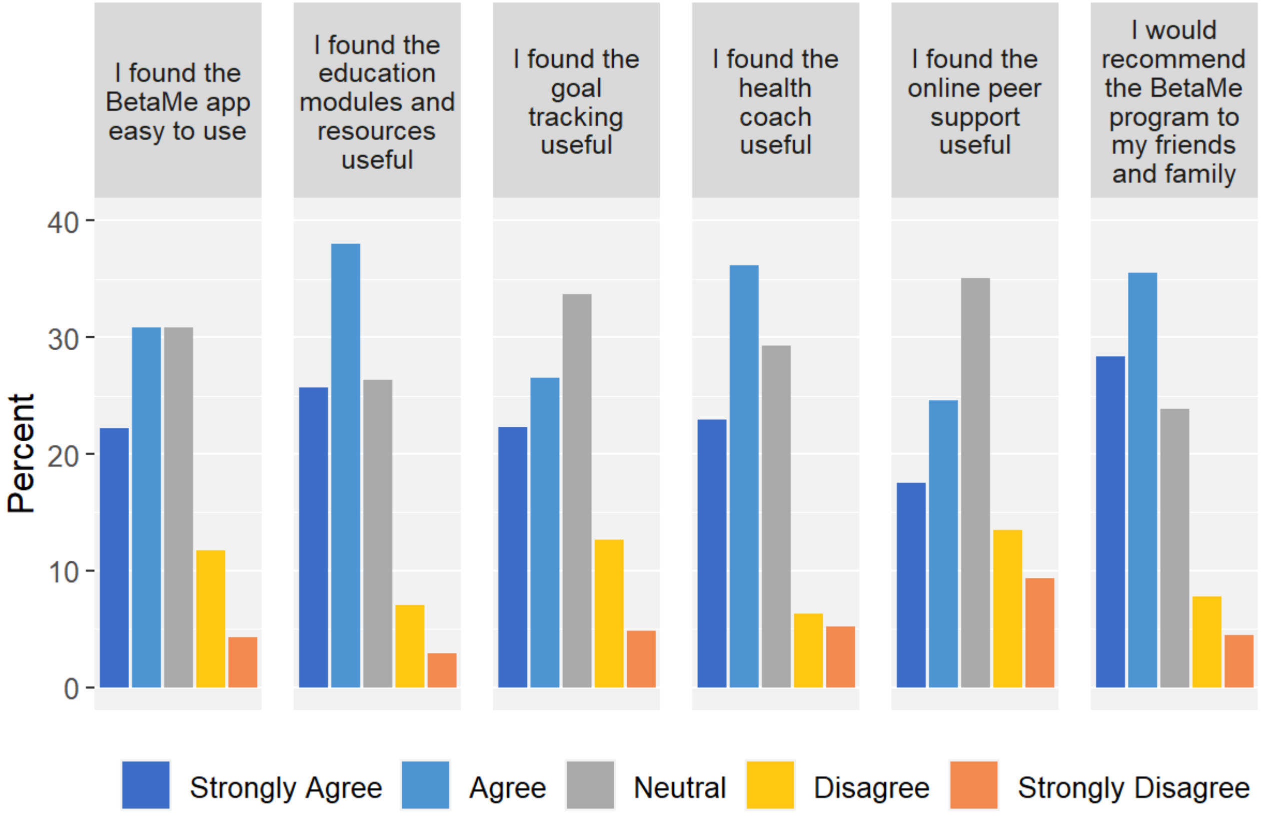


Participant Perceptions

Online Survey

The online questionnaire on the usefulness of the entire BetaMe/Melon program (Figure 4) was completed by 83.7% (180/215) intervention arm participants during the final RCT assessment at 12 months. However, not all participants completed all items. Overall, participants rated the education resources and health coaches as more useful components of the program (63.7% [109/171] and 59.2% [103/174] "agreed" or "strongly agreed" with these statements, respectively). Participants considered goal tracking and the online peer support less useful (48.8% [81/166] and 42.1% [72/171] "agreed" or "strongly agreed" with these statements, respectively). About half of respondents (53.0%, 86/162) "agreed" or "strongly agreed" that the program was easy to use. The majority of participants (63.9%, 115/180) indicated that they would recommend the program to friends and family.

Figure 4. Individual participant feedback from the online questionnaire (percent giving specific response to each item).



Participant Interviews

Of the 18 participants who participated in the telephone interviews, 10 were female. Most participants were aged between 45 and 74 years (n=1 aged 35-44, n=10 aged 45-64, n=7 aged 65-74). The majority (n=10) were of non-Pacific/non-Māori ethnicity, 5 were Māori, and 3 were Pacific. Eleven of the 18 participants had HbA_{1c} levels in the diabetes range and 7 had HbA_{1c} levels in the prediabetes range. Interviews lasted 10 to 30 minutes (mean=18 minutes). Thematic analysis of the interviews resulted in four themes about the 12-month program, each of which are described in the following sections and exemplified in Table 2.

Table 2. Coding framework extract with example categories and codes.

Theme	Example codes
Overall program utility	
Psychological experience	Good and helpful, helped establish change, improved coping with diabetes, liked being monitored, disliked misinformation about program, lack of attention led to regression
Technological experience	Program easy to access, trouble with video conferencing, easy, could not access component
Physical experience	Reduced amount of food eaten, made small changes
Health coach	
Psychological experience	Bubbly, helpful, responsive, easy to understand, supportive, shared personal experiences, motivating, related to coach of same ethnicity, good reinforcement.
Suggested improvements	Coach continuity, increased coach contact, coaches that are role models (eg, same gender, similar age), continuity of contact, improved follow-up
Least useful components	
Online peer support use	Did not use, did not interact, did not know about this component, read some, not needed, could not access
Online peer support psychological experience	Not comfortable with medium, did not like, worked well, influential, could not relate to others
Goal tracking use	Did own tracking, set own goals, easy, did not know about it, did not use it
Goal tracking psychological experience	Good, helpful, useful, loved it, unsure, unenthusiastic

Overall Program Utility

Interview participants found the BetaMe/Melon program to be useful overall, with most indicating that they had a positive experience and identifying beneficial outcomes from participating. For example, a number reported that the program increased their knowledge and led to behavioral changes that helped them lose weight and better manage their diabetes or prediabetes.

> *Being made more aware of my eating habits. Because being Māori Chinese, when we were little kids we were always told, don't waste your food, don't waste your food, so therefore I kept on eating my food, eating my food. And I had a whole sort of regime where I would eat everything on my plate, don't waste the food.* [Participant #10]

Most participants reported feeling motivated when they commenced the program and were able to identify components of the program that helped to maintain their motivation. These components included regular contact with the health coach, medication reminders, and the ability to track results relevant to their goals (eg, blood sugar regulation, weight management, and exercising).

> *The things I did find useful were the reminders to take your medication, I found them very helpful especially the night one cause I wasn't used to taking that.* [Participant #13]

The majority of participants reported that they would recommend the program to others.

Participant Feedback Regarding Health Coach Element

Connecting with the health coach was the most useful component identified. Participants articulated that regular contact with the online health coaches gave them someone to be accountable to and helped them remain goal oriented. The health coaches were reported as being good communicators, encouraging, positive, and willing to connect with participants on a personal level. One participant indicated that they were so happy with their coach that they wanted to maintain contact following the 16-week active support component of the program.

> *Very good, very helpful, always there if you needed advice, they were always on the end of the email, so from that point of view they were very, very effective.* [Participant #4]

However, a number of participants had difficulties connecting or staying connected to their health coach, with one participant explicitly stating they would have preferred a relatable health coach, such as a coach who had been through a weight loss journey and thus would be a role model to aspire to.

> *Well I'd prefer a female, and someone around my… age I can relate to, or ish, who's been there and done that, that I can like sort of aspire to like, yay she's done it maybe I can do it to.* [Participant #18]

Online Peer Support and Goal Tracking

Participants found the online peer support and goal tracking to be the least useful components. Over half of interviewed participants did not use the online peer support component, with personal preference being the primary reason. One participant had technical difficulties accessing the peer support and another reported not knowing about the peer support component of BetaMe/Melon. Three participants read forum posts from time to time but interacted minimally online. Another stopped using the peer support forum, stating that she was unable to relate to other participants due to differences in affordable food options, as what they talked about did not match her budget.

> *I did put in a comment, you know, has anybody got ideas with pigs head, because that's just the type of meat I can afford.…It wasn't that beneficial for me, because I was looking at cheap meats, and they were looking at meats I don't look at.…I felt like I was the poor one in the group because my questions were so to the left of what everyone else was on.* [Participant #16]

The goal tracking component was used by 7 of the interviewed participants, all of whom stated that they found it helpful. Another 5 participants tracked progress against their goals but did not use the BetaMe/Melon program to do so. Six participants did not track their progress against goals at all, although this group included 2 participants who were not aware of the goal tracking component in the program.

> *I kept getting reminders saying would you input your data, you know your blood sugars, but I didn't know how to do that or what to do so I never did it. But I keep my own records you know.* [Participant #13]

Barriers and Recommendations

Participants identified a range of problems with the program as well as barriers to its use in daily life. These included problems with the functionality of the program or with internet connectivity; incomplete delivery of all program components for some participants (eg, not setting goals with the health coach or not being informed about all components); and barriers related to the skills, knowledge, or motivation of participants.

Participants suggested a range of improvements to the BetaMe/Melon Program. Suggestions included better support and training regarding how to load and use each component of the program, functionality improvements (such as enabling participants to record specific types of exercise undertaken or share goal tracking information with other willing participants), more frequent and longer contact with the health coach, face-to-face rather than online health coaching, keeping the same coach throughout the program, and use of coaches who are relatable role models. The need to recognize that each participant was unique was commonly suggested. For example, participants suggested efficacy would be improved if BetaMe/Melon were to consider participants' prior experiences and relative states of minds when commencing the program, including participants' experiences with goal tracking and weight loss programs, their levels of motivation, and known enablers and barriers to self-management of health conditions. Other suggestions for individualization included increasing program flexibility in order to adapt to participants' changing life circumstances (eg, being on holiday, unwell, or hospitalized) and tailoring program information and resources to individual participants (eg, providing recipes to cater for those with limited resources).

Suggestions regarding the online peer support included considering other ways of providing this component such as connecting participants so they could meet to share experiences, motivate each other, and participate in group exercise or walks. Participants also suggested that family be included in the program and that a counselling feature be implemented. Finally, participants recommended that the program provide advice on where to go to for help upon completion of the program.

Discussion

Overview

This process evaluation was part of a wider study including an RCT [15,26]. It utilized data from 3 sources to assess engagement with and usefulness of a comprehensive digital health intervention program. Overall, 92% (198/215) of intervention participants completed the compulsory first health coach session, although engagement with other components of the program varied by age, ethnicity, and gender and fell over time. Despite this, 64% of respondents (115/180) of participants indicated that they would recommend the program to others, and the majority of those taking part in qualitative interviews following the study period indicated that they found the program to be useful overall.

Principal Findings

Engagement with the health coach for the first nonoptional coaching session was high at 92% (198/215). Active engagement with other online components of the program peaked in Week 0 with 50% (108/215) of participants actively engaging online. Similar to other programs, participant engagement with the BetaMe/Melon digital health intervention program varied by demographic, with women, younger participants, and ethnic majority population groups demonstrating higher engagement [27,28]. However, engagement dropped steadily over time [28-30]. Reduced engagement over time is a consistent issue with digital health intervention programs [28] and is likely to relate to the attenuation of any health gains achieved by an intervention over time [27,31]. Improving the design of digital health intervention programs to maintain and increase engagement (eg, offering regular rewards, introducing new content to sustain interest, or emailing engagement reminders) and involving specific groups for which the program is likely to be less successful at engaging (either initially or over time) in program development may improve engagement and intervention success [12,28].

We found lower overall active engagement with the online components of the program for Māori and Pacific participants. The BetaMe/Melon program was developed using behavioral change and cognitive behavioral theory [30] and included input from Māori and Pacific health care providers. Given the higher prevalence of diabetes and prediabetes in the Māori and Pacific populations of New Zealand, it may be useful for the BetaMe/Melon program (or other digital health intervention programs in New Zealand) to consider using relevant models and behavior change theory specific to Māori and Pacific peoples [32,33]. Internationally, focusing on ethnically appropriate theory when developing and refining digital health interventions may improve engagement and impact for indigenous and minority population groups [31,33,34].

Multimodal programs like BetaMe/Melon have been shown in some studies to be more successful than those adopting a single mode [9,21], with other studies showing that the success of multimodal digital health intervention programs can be moderated by the impact of diabetes self-management style on engagement with and utilization of various components [35]. In general, the BetaMe/Melon program was well received, with educational resources and health coaching seen as the most useful components. However, there was a strong call by participants to strengthen the human components of the program or, in other words, to implement both longer or face-to-face contact with the health coach as well as enhanced peer support activities. Conversely, goal tracking was identified as the least useful component of BetaMe/Melon by participants in both the online survey and telephone interviews. However, diary entries to track progress toward health goals were the most utilized form of engagement irrespective of participant characteristics. Additionally, it appears that some participants were unaware of all components of the BetaMe/Melon program (goal tracking and peer support forum). A few barriers to incorporating the program in everyday life were identified, and several improvements proposed. In particular, the need for increased individualization of the program, a bigger role for the health coaches, and technical improvements to the online components of the program to improve usability were emphasized. Other studies have also called for digital health intervention programs to be tailored to individual participants' needs [9,12,28,35], to combine digital and human support in order to increase engagement [12,36], and particularly to enable feedback loops between the participant and their health care team in order to improve diabetes control [1,9,12].

Strengths and Limitations

A key strength of this study is that it was conducted in the context of an RCT of the comprehensive BetaMe/Melon digital health intervention program. We have addressed knowledge gaps around comprehensive self-management programs for people with diabetes and prediabetes and for Māori and Pacific peoples. Another strength is that we were able to describe utilization over time and by type of activity and demographic characteristics (age, ethnicity, gender, and diabetic status).

This study also has some limitations. First, in assessing online engagement with the study, we were limited to the data released by Melon Health Ltd to the Otago University study team. From the data that were provided, we were able to identify the number of completed first health coaching sessions and active engagement with the online program (diary entries and use of the peer support forum). Melon Health Ltd did not supply data on the number or length of health coach sessions per participant or other online activity such as login activity or passive engagement with the program (eg, reading posts or accessing educational material). Second, not all eligible participants answered the online questionnaire, and the participants that agreed to talk with us following the completion of the program may not be representative of the entire study population. Finally, participants were asked 12 months after enrollment to recall their perceptions of a program with a primary focus on the first 16 weeks of delivery. Thus, there is potential for inaccurate recall by participants, and generalizing the findings beyond this study may need careful consideration.

Conclusions

The BetaMe/Melon program was generally well received by participants, and initial engagement with the health coach was high, although engagement with other components varied by participant demographic and dropped rapidly over the 16-week active program period. Maintaining engagement of participants over time, individualizing programs, and addressing technical barriers to engagement are important to address in order to maximize satisfaction, engagement, and potential health benefits that may come from digital programs for self-management of diabetes.

Acknowledgments

This study was funded by the Health Research Council of New Zealand, the Ministry of Health New Zealand, and the Healthier Lives National Science Challenge, 16/727. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or the writing of this paper.

Authors' Contributions

Study concept and design was led by DS, MM, and J Stanley with input from J Stairmand, VS, JK, RG, and AD. The study was managed by MM. MM, J Stanley, and DS led the analysis and interpretation of data overall. The utilization data was acquired by MM and analyzed and interpreted by MM, J Stanley, and NS. The online survey data was acquired by VS, DT, and KH and analyzed and interpreted by MM, J Stanley, and VS. The patient interview data was acquired by J Stairmand, analyzed and interpreted by J Stairmand and VS with the support of MM and CD. VS, MM, and J Stanley wrote the first draft of the manuscript with input from J Stairmand and DS. VS, MM, J Stanley, J Stairmand, DS, NS, DT, KH, CD, JK, AD, and RG provided critical revision of the manuscript draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
BetaMe/Melon Process Evaluation Telephone Interview Schedule.
[PDF File (Adobe PDF File), 286 KB - jmir_v22i12e19150_app1.pdf]

References

1. van Dieren S, Beulens JW, van der Schouw YT, Grobbee DE, Neal B. The global burden of diabetes and its complications: an emerging pandemic. Eur J Cardiovasc Prev Rehabil 2010 May;17 Suppl 1:S3-S8. [doi: 10.1097/01.hjr.0000368191.86614.5a] [Medline: 20489418]
2. Hajat C, Stein E. The global burden of multiple chronic conditions: A narrative review. Prev Med Rep 2018 Dec;12:284-293 [FREE Full text] [doi: 10.1016/j.pmedr.2018.10.008] [Medline: 30406006]
3. Ministry of Health. Health Loss in New Zealand: A report from the New Zealand Burden of Diseases, Injuries and Risk Factors Study, 2006–2016. Wellington: Ministry of Health; 2013. URL: https://www.moh.govt.nz/notebook/nbbooks.nsf/0/F85C39E4495B9684CC257BD3006F6299/$file/health-loss-in-new-zealand-final.pdf [accessed 2020-11-24]
4. Ministry of Health. Living Well with Diabetes: A plan for people at high risk of or living with diabetes 2015–2020. Ministry of Health. Wellington: Ministry of Health; 2015 Oct 16. URL: https://www.health.govt.nz/system/files/documents/publications/living-well-with-diabetes-oct15.pdf [accessed 2020-11-24]
5. Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res Clin Pract 2011 Dec;94(3):311-321. [doi: 10.1016/j.diabres.2011.10.029] [Medline: 22079683]
6. Bhutani J, Bhutani S. Worldwide burden of diabetes. Indian J Endocrinol Metab 2014 Nov;18(6):868-870 [FREE Full text] [doi: 10.4103/2230-8210.141388] [Medline: 25364686]
7. Reynolds R, Dennis S, Hasan I, Slewa J, Chen W, Tian D, et al. A systematic review of chronic disease management interventions in primary care. BMC Fam Pract 2018 Jan 09;19(1):11 [FREE Full text] [doi: 10.1186/s12875-017-0692-3] [Medline: 29316889]
8. Sepah SC, Jiang L, Peters AL. Long-term outcomes of a Web-based diabetes prevention program: 2-year results of a single-arm longitudinal study. J Med Internet Res 2015 Apr 10;17(4):e92 [FREE Full text] [doi: 10.2196/jmir.4052] [Medline: 25863515]

9. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. J Diabetes Sci Technol 2017 Sep 31;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]
10. Captieux M, Pearce G, Parke HL, Epiphaniou E, Wild S, Taylor SJC, et al. Supported self-management for people with type 2 diabetes: a meta-review of quantitative systematic reviews. BMJ Open 2018 Dec 14;8(12):e024262 [FREE Full text] [doi: 10.1136/bmjopen-2018-024262] [Medline: 30552277]
11. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
12. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
13. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
14. Klonoff DC. Behavioral Theory: The Missing Ingredient for Digital Health Tools to Change Behavior and Increase Adherence. J Diabetes Sci Technol 2019 Mar 24;13(2):276-281 [FREE Full text] [doi: 10.1177/1932296818820303] [Medline: 30678472]
15. Sarfati D, McLeod M, Stanley J, Signal V, Stairmand J, Krebs J, et al. BetaMe: impact of a comprehensive digital health programme on HbA1c and weight at 12 months for people with diabetes and pre-diabetes: study protocol for a randomised controlled trial. Trials 2018 Mar 05;19(1):161 [FREE Full text] [doi: 10.1186/s13063-018-2528-4] [Medline: 29506562]
16. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A Mobile Health Intervention for Self-Management and Lifestyle Change for Persons With Type 2 Diabetes, Part 2: One-Year Results From the Norwegian Randomized Controlled Trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
17. Ristau RA, Yang J, White JR. Evaluation and Evolution of Diabetes Mobile Applications: Key Factors for Health Care Professionals Seeking to Guide Patients. Diabetes Spectrum 2013 Nov 15;26(4):211-215. [doi: 10.2337/diaspect.26.4.211]
18. Cho J, Lee H, Lim D, Kwon H, Yoon K. Mobile communication using a mobile phone with a glucometer for glucose control in Type 2 patients with diabetes: as effective as an Internet-based glucose monitoring system. J Telemed Telecare 2009 Mar;15(2):77-82. [doi: 10.1258/jtt.2008.080412] [Medline: 19246607]
19. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011 Apr;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]
20. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic Health Conditions and Internet Behavioral Interventions. CIN: Computers, Informatics, Nursing 2011;29(2):81-92. [doi: 10.1097/ncn.0b013e3182065eed] [Medline: 6424131]
21. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun 8;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
22. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
23. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: UK Medical Research Council (MRC) guidance. URL: https://mrc.ukri.org/documents/pdf/mrc-phsrn-process-evaluation-guidance-final/ [accessed 2020-11-24]
24. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 2700030]
25. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. Journal of Biomedical Informatics 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208]
26. McLeod M, Stanley J, Signal V, Stairmand J, Thompson D, Henderson K, et al. Impact of a comprehensive digital health programme on HbA and weight after 12 months for people with diabetes and prediabetes: a randomised controlled trial. Diabetologia 2020 Dec 04;63(12):2559-2570. [doi: 10.1007/s00125-020-05261-x] [Medline: 32886192]
27. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
28. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of User Engagement with Mobile- and Web-Delivered Self-Care Interventions for Adults with T2DM: A Review of the Literature. Curr Diab Rep 2016 Jul 2;16(7):66 [FREE Full text] [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]

29. Geraghty AWA, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international Internet health interventions: a step towards global eHealth. Health Promot Int 2013 Sep 10;28(3):442-452 [FREE Full text] [doi: 10.1093/heapro/das029] [Medline: 22786673]
30. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
31. McSharry J, Byrne M, Casey B, Dinneen SF, Fredrix M, Hynes L, et al. Behaviour change in diabetes: behavioural science advancements to support the use of theory. Diabet Med 2020 Mar 04;37(3):455-463. [doi: 10.1111/dme.14198] [Medline: 31797455]
32. Beaton A, Manuel C, Tapsell J, Foote J, Oetzel JG, Hudson M. He Pikinga Waiora: supporting Māori health organisations to respond to pre-diabetes. Int J Equity Health 2019 Jan 07;18(1):3 [FREE Full text] [doi: 10.1186/s12939-018-0904-z] [Medline: 30612567]
33. Oetzel J, Scott N, Hudson M, Masters B, Rarere M, Foote J, et al. He Pikinga Waiora Implementation Framework: A tool for chronic disease intervention effectiveness in Māori and other indigenous communities. Int J Integr Care 2018 Mar 12;18(s1):68. [doi: 10.5334/ijic.s1068]
34. Bender MS, Cooper BA, Park LG, Padash S, Arai S. A Feasible and Efficacious Mobile-Phone Based Lifestyle Intervention for Filipino Americans with Type 2 Diabetes: Randomized Controlled Trial. JMIR Diabetes 2017 Dec 12;2(2):e30. [doi: 10.2196/diabetes.8156]
35. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. User Experiences With a Type 2 Diabetes Coaching App: Qualitative Study. JMIR Diabetes 2020 Jul 17;5(3):e16692 [FREE Full text] [doi: 10.2196/16692] [Medline: 32706649]
36. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout From an eHealth Intervention for Adults With Type 2 Diabetes: A Qualitative Study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]

Abbreviations

HbA_{1c}: hemoglobin A_{1c}
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 06.04.20; peer-reviewed by X Ji, S Wakamiya, M Fajardo, E Pustozerov; comments to author 12.06.20; revised version received 15.10.20; accepted 26.10.20; published 01.12.20.

Please cite as:
Signal V, McLeod M, Stanley J, Stairmand J, Sukumaran N, Thompson DM, Henderson K, Davies C, Krebs J, Dowell A, Grainger R, Sarfati D
A Mobile- and Web-Based Health Intervention Program for Diabetes and Prediabetes Self-Management (BetaMe/Melon): Process Evaluation Following a Randomized Controlled Trial
J Med Internet Res 2020;22(12):e19150
URL: https://www.jmir.org/2020/12/e19150
doi:10.2196/19150
PMID:33258776

Original Paper

Effectiveness of WeChat for Improving Exclusive Breastfeeding in Huzhu County China: Randomized Controlled Trial

Qiong Wu[1], MSc; Yiwen Huang[1], MSc; Zijun Liao[1], PhD; Michelle Helena van Velthoven[2], PhD; Wei Wang[1], MSc; Yanfeng Zhang[1], MSc

[1]Capital Institute of Pediatrics, Beijing, China
[2]University of Oxford, Oxford, United Kingdom

Corresponding Author:
Yanfeng Zhang, MSc
Capital Institute of Pediatrics
No. 2 Yabao Road
Chaoyang District
Beijing, 100020
China
Phone: 86 10 85695554
Fax: 86 10 85622025
Email: summyzh@126.com

Abstract

Background: The benefits of breastfeeding for both infants and mothers have been well recognized. However, the exclusive breastfeeding rate in China is low and decreasing. Mobile technologies have rapidly developed; communication apps such as WeChat (one of the largest social networking platforms in China) are widely used and have the potential to conveniently improve health behaviors.

Objective: This study aimed to assess the effectiveness of using WeChat to improve breastfeeding practices.

Methods: This 2-arm randomized controlled trial was conducted among pregnant women from May 2019 to April 2020 in Huzhu County, Qinghai Province, China. Pregnant women were eligible to participate if they were aged 18 years or older, were 11 to 37 weeks pregnant with a singleton fetus, had no known illness that could limit breastfeeding after childbirth, used WeChat through their smartphone, and had access to the internet. A total of 344 pregnant women were recruited at baseline, with 170 in the intervention group and 174 in the control group. Women in the intervention group received breastfeeding knowledge and promotion information weekly through a WeChat official account from their third month of pregnancy to 6 months postpartum. The primary outcome of exclusive and predominant breastfeeding rate was measured 0-1 month, 2-3 months, and 4-5 months postpartum.

Results: At 0-1 month postpartum, the exclusive breastfeeding rate was significantly higher in the intervention group than that in the control group (81.1% vs 63.3%; odds ratio [OR] 2.75, 95% CI 1.58-4.78; *P*<.001). Similarly, mothers in the intervention group were more likely to provide predominantly breast milk (OR 2.77, 95% CI 1.55-4.96; *P*<.001) and less likely to give dairy products to their children (OR 0.40, 95% CI 0.21-0.75; *P*=.005). There was no statistically significant difference for exclusive breastfeeding rate 2-3 months (*P*=.09) and 4-5 months postpartum (*P*=.27), though more children in the intervention group were exclusively breastfed than those in the control group 2-3 months postpartum (intervention: 111/152, 73.0%; control: 96/152, 63.2%) and 4-5 months postpartum(intervention: 50/108, 46.3%; control: 46/109, 42.2%).

Conclusions: This study is the first effort to promote exclusive breastfeeding through WeChat in China, which proved to be an effective method of promoting exclusive breastfeeding in early life. WeChat health education can be used in addition to local breastfeeding promotion programs.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1800017364; http://www.chictr.org.cn/showproj.aspx?proj=29325

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-019-7676-2

(J Med Internet Res 2020;22(12):e23273) doi:10.2196/23273

KEYWORDS

breastfeeding; exclusive breastfeeding; WeChat; mHealth; randomized controlled trial

Introduction

Appropriate child feeding is the foundation for good nutritional intake and healthy development and is a critical factor for health in adults [1-3]. As a part of optimal feeding practices, exclusive breastfeeding is recognized as a cornerstone of child survival and health, by providing essential irreplaceable nutrition for a child's growth and development [4]. Therefore, the World Health Organization (WHO) and United Nations Children's Fund (UNICEF) recommend that children should be exclusively breastfed from birth to 6 months of age and continually breastfed until they are 2 years old or older [5]. Moreover, one of the WHO global nutrition targets for 2025 is increasing the exclusive breastfeeding rate in the first 6 months of life to at least 50% [4].

In China, the exclusive breastfeeding rate for under 6 months of age is a concern and achieving the 2025 global target remains an ongoing challenge. According to the Chinese national data, the exclusive breastfeeding rate under 6 months was 27.6% in 2008 [6] and 20.7% in 2013 (the weighted exclusive breastfeeding rate was 18.6%) [7], which showed a downward trend. Therefore, much effort is required to explore effective ways to control this downward trend, and hence, to promote breastfeeding in China.

Positive breastfeeding outcomes are usually related to the improvement of maternal breastfeeding knowledge and attitudes [2,8]. Although interventions on breastfeeding promotion vary worldwide, education and support are the 2 most common approaches, and interventions on exclusive breastfeeding containing these 2 elements have mixed success [9,10]. In China, breastfeeding education is generally implemented through the rural 3-tier health care system (county-township-village) at antenatal and postnatal care [11]. Since 2009, China has been implementing a national program called Basic Public Health Service, in which health care workers are required to provide face-to-face breastfeeding and complementary feeding counseling to pregnant women and mothers throughout antenatal and postnatal care [11]. However, research has indicated that mothers in rural areas rarely receive feeding information from health facilities; their main sources of information were family members and friends, who were unlikely to have access to better information and may have misinformed mothers [12-14]. Therefore, new channels are needed in rural China to deliver effective infant feeding education.

With the widespread use of smartphones, using apps in the health sector for delivering health care services and health promotion is an increasing phenomenon [15-19]. In China, one of the most popular smartphone apps is WeChat, which offers services such as Facebook, Twitter, WhatsApp, and others, on a single platform. More than 1.32 billion users were registered with WeChat throughout the world by Q2 2019, and more than 1.15 billion people were monthly active users [20]. Approximately 45 billion messages are exchanged on the platform every day [20]. Furthermore, there is an app-within-an-app platform in WeChat called WeChat Official Accounts, which can be used for individuals, governments, media organizations, and business enterprises to communicate and interact with their subscribers and provide them with services through text, images, voice, videos, and rich-media messages [21]. There were more than 20 million registered WeChat official accounts at the end of 2018 [20]. WeChat is gradually changing the channels through which people receive information and has been used as a communication tool to change health behaviors, showing potential positive impacts on disease management of cancer [22], malaria [23], asthma [24], chronic rhinosinusitis [25], diabetes [26], and weight loss [27]. However, no studies have focused on using WeChat to support caregivers with infant and young child feeding. The objective of this randomized controlled trial was to evaluate the effectiveness of a WeChat breastfeeding intervention on promoting exclusive breastfeeding in rural areas of China.

Methods

Study Design

A 2-arm randomized controlled trial (Chinese Clinical Trial Registry ChiCTR1800017364) was conducted between May 9, 2019 and April 3, 2020. We aimed to evaluate the effectiveness of using a WeChat account for improving exclusive breastfeeding of children aged 0-5 months. The protocol for this randomized controlled trial was previously published [28]. The sampling unit was individual pregnant women, who were randomized to routine antenatal and postnatal care or routine care plus the WeChat breastfeeding education.

Study Sites and Context

This trial was carried out in 13 townships in Huzhu County, Qinghai Province, China. Qinghai Province is in northwest China, with a total population of 6,078,200 in 2019. Huzhu County lies in the northeast of Qinghai province and has a total population of 401,540 [29]. There were 91,321 women of reproductive age and 4325 pregnant women in Huzhu County in 2017 (Huzhu County Maternal and Child Health Family Planning Service Centre).

Huzhu County has 19 townships and 294 villages. We excluded 6 townships; 4 townships had already been selected by another maternal and child health project, and the other 2 were remote with a small number of pregnant women.

Participants and Recruitment

Pregnant women were eligible to participate in this trial if they were aged at least 18 years, were 11-37 weeks pregnant with a singleton fetus, had no known illness that limits breastfeeding after childbirth, were able to read and communicate in Mandarin, used WeChat through their smartphone, and had access to the internet. The exclusion criteria were (1) pregnant women who did not come to the township hospitals to participate in the trial; (2) pregnant women with severe disease and complications of pregnancy or HIV; (3) women who had a miscarriage or stillbirth; (4) mothers with infants with a low birth weight

(<2500 g) or who were born prematurely (<37 weeks of gestation).

Before recruitment, we asked each township hospital to provide a list of the names of all pregnant women between 11 to 37 weeks' gestation, which included information on gestational age, gravidity, and parity. A total of 444 women were listed. Based on this list, participating pregnant women were invited to come to their township hospitals and were given full information about the study between May 9, 2019 and May 17, 2019. After agreeing to participate and signing the written consent form, a researcher gave each eligible pregnant woman an opaque sealed envelope, which included a random number generated in advance and indicated the allocated group. After completing baseline data collection, all the eligible participants were randomized to either the WeChat intervention group or the control group at a 1:1 ratio.

Sample Size Calculation

The sample size for this study was estimated using a 0-5 months of age exclusive breastfeeding rate of 29.2% from a pilot cross-sectional survey conducted by our team in Datong County in Qinghai in September 2017. We expected to achieve a 20% increase in the exclusive breastfeeding rate with the WeChat intervention. Assuming a power of 80% and a 5% significance level, we determined that a sample size of 93 pregnant women for each (intervention and control) group was needed for this study. To compensate for attrition and loss to follow-up, we planned to enroll 200 pregnant women in each group.

Intervention Group

Women in both intervention and control groups were asked to follow the WeChat account called Huzhu County Maternal and Child Health Family Planning Service Centre on their smartphone by scanning the 2D code at the back of random number cards. There was a special module called *Ke Xue Wei Yang* (*Optimal Feeding*) within the WeChat Official Account which was developed by an information technology company, ZYZY (Beijing) Pioneer of Cultural Essence Co Ltd and pretested in Huzhu County in Aug 2018 [28]. Pregnant women allocated to the intervention group were also asked to subscribe to and register with the *Ke Xue Wei Yang* module by entering information on their name, phone number, gestational age, expected date of delivery, village and county of residence and were provided a WeChat log-in and password. Pregnant women in the control group were not able to register with the *Ke Xue Wei Yang* module and did not have access to the information in the module, to preclude contamination between groups through direct sharing of messages sent via WeChat.

There were 4 components in the *Ke Xue Wei Yang* module: feeding messages, a feeding knowledge competition, a baby growth chart, and an online forum (Figure 1), which were described in detail in our protocol [28].

The feeding messages, the most important component, were designed for breastfeeding promotion education and provided key breastfeeding knowledge and relevant infant feeding advice, breastfeeding problems encountered for both mother and child, and preparation for both breastfeeding and complementary feeding. All messages were developed based on the WHO breastfeeding recommendations, guidelines, or published literature; messages were published before recruitment. Once women in the intervention group subscribed to the *Ke Xue Wei Yang* module, they could read all messages whenever they wanted. Furthermore, we considered that late pregnancy (37 weeks or above), the first month postpartum, and 4 months postpartum were 3 key stages for mothers and that breastfeeding information needed to be strengthened during these stages, so we sent an additional 3 sets of tailored messages to each pregnant woman and mother at these stages via WeChat on Monday, Wednesday, and Friday every week. Specifically, we sent information on getting ready for breastfeeding and key breastfeeding recommendations to women who were at least 37 weeks pregnant. For new mothers 1 month postpartum, we sent key breastfeeding recommendations and common breastfeeding problems encountered for both mother and child. We sent information on starting complementary feeding by 6 months of age to mothers whose children were 4 months or older (to avoid having mothers introduce complementary food too early or too late).

Women could participate in the feeding knowledge competition component to test their breastfeeding knowledge. Moreover, women could enter their children's weight and height in the baby growth chart component whenever they want to monitor their children's growth and ask breastfeeding related questions on the online forum component.

Figure 1. The interface of the WeChat intervention: (a) log-in interface, (b) main interface, (c) feeding messages, (d) feeding competition, (e) baby growth chart, and (f) online forum.

图上有7条曲线，由上而下分别代表同年龄宝宝分布的第99、97、85、50、15、3、1百分位线。红点为该月龄下宝宝的身高位置。一般而言，婴儿生长指标若落在第97及第3百分位两线之间均属正常，否则要考虑宝宝的身高指标有过高或过低的情况。

亲，您可以在社区中发帖提问哦

帖子7 - 浏览55 - 回复5

wx_叶苏_JkChE 08-03 13:47

【本社区已开通专家提问答复服务】

阅读 29 回复 1

wx_艺文_6ISIL 04-15 14:59

宝宝晚上吃几次奶合适啊

阅读 15 回复 2

still water 04-15 14:58

我要提问

阅读 0 回复 1

Data Collection

Data were collected by face-to-face at baseline and by telephone during follow-up interviews. Baseline data collection was carried out after the recruitment session. After receiving consent to participate, we interviewed the pregnant women (baseline questionnaire), which included information on demographic characteristics, antenatal care, hemoglobin levels, as well as breastfeeding knowledge. Baseline interviewers, who were from a vocational-technical school in Huzhu County and trained on interview methodology, used smartphones to collect the baseline data.

Follow-up interviews were conducted by telephone at 0-1 months (mean 35.3 days, SD 10.5), 2-3 months (mean 102.7 days, SD 9.9), 4-5 months (mean 157.2 days, SD 17.8) postpartum. Data from women in both the intervention group and the control group were collected on breastfeeding knowledge, practices, reasons for weaning, and information channels by health workers form Huzhu County Maternal and Child Health Family Planning Service Centre, who were blinded to intervention status and trained on interview methodology by research staff from the Capital Institute of Pediatrics.

Outcome Measurement

The primary outcome measure was exclusive and predominant breastfeeding rate at 0-1 month (0-60 days), 2-3 months (61-120 days), and 4-5 months (121-180 days) postpartum in both the intervention group and the control group. Secondary outcomes included the following: (1) the proportion of early initiation of breastfeeding; (2) prelacteal feeding rate; (3) rate of any breastfeeding; (4) mothers' knowledge on breastfeeding practices; (5) other infant feeding practices (such as giving dairy or dairy products, water, semisolid, or solid foods at 3 follow-ups).

According to the WHO definition, *exclusive breastfeeding* is defined as an infant receiving only breast milk, no other liquids or solids, except oral rehydration solution, drops or syrups for vitamins, minerals supplementation, or medicine [30]; *predominant breastfeeding* permits partial substitution with water-based fluids; and *early initiation of breastfeeding* is when an infant is put to the mother's breast within 1 hour of birth [30]. *Prelacteal feeding* was defined as the newborn being provided any food except mother's milk before initiating breastfeeding; *any breastfeeding* included partial substitution with infant formula, other fluids, or solid foods.

Data Management and Analysis

We performed statistical analysis with SAS (version 9.2 for Windows; SAS Institute). We summarized baseline characteristics, follow-up infant feeding practices, and mother's infant feeding knowledge, as median and interquartile range (IQR) for continuous variables or as number and proportion for categorical variables. We estimated the homogeneity in baseline characteristics between groups using the Wilcoxon rank-sum test for nonparametric continuous variables, and chi-square or Fisher exact test for categorical variables. Multiple logistic regression models were used to estimate the effects of the intervention on breastfeeding rates between groups at the 3 follow-up points, controlling for parity at baseline, as the distribution of parity was different between the 2 groups. We also used chi-square tests to compare the differences in mother's infant feeding knowledge between groups at baseline and 3 follow-up points. Participants who dropped out or who were lost to follow-up during the postpartum period were excluded from analysis and are reported separately (Multimedia Appendix 1). A *P* value <.05 was considered statistically significant.

Ethics Statement

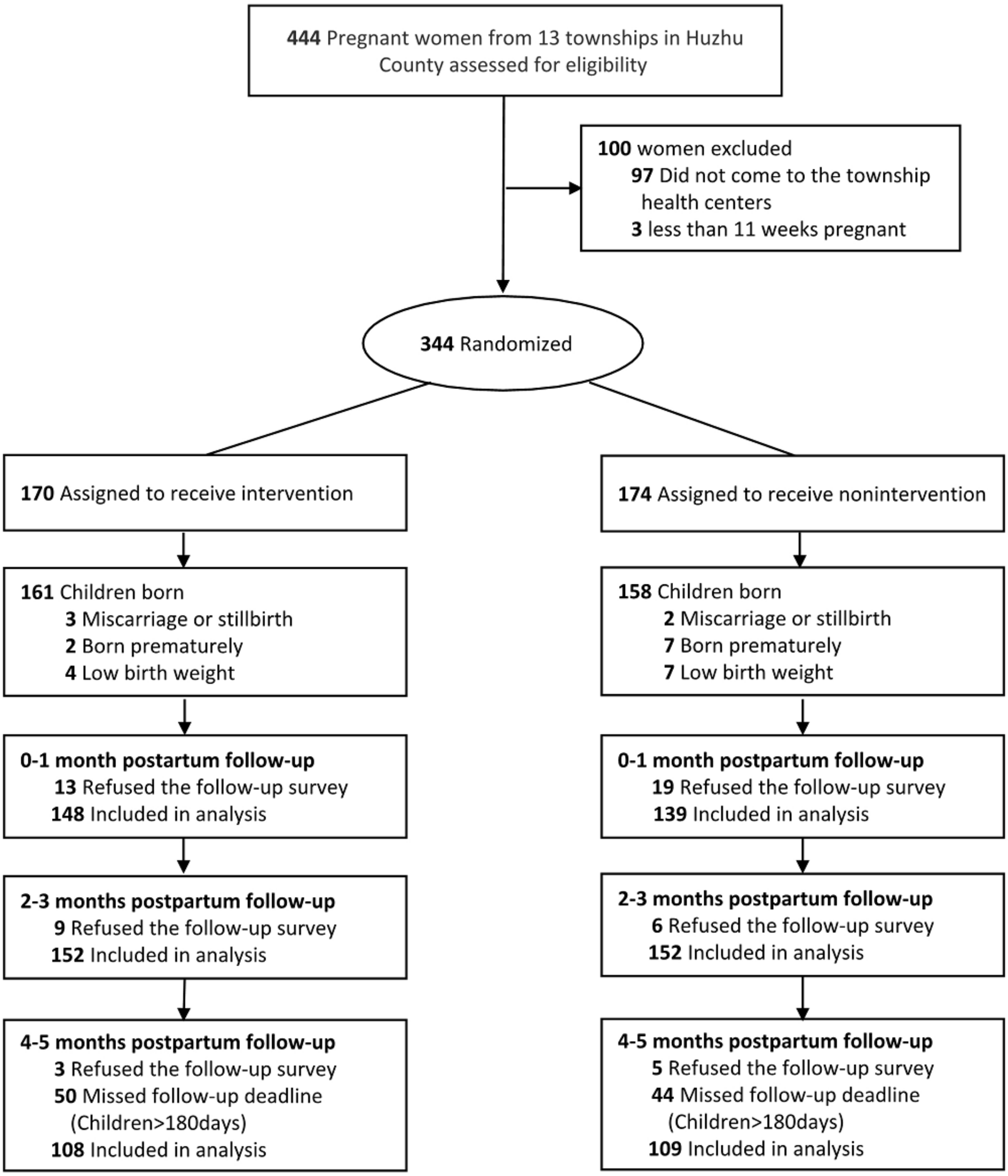
This study was approved by the ethical committee of the Capital Institute of Pediatrics, Beijing, China. All women were given an information sheet, and participating women provided both verbal and written informed consent.

Results

General

We recruited 344 pregnant women. We randomized 170 to the intervention group and 174 to the control group between May 9, 2019 and May 17, 2019; 25 were excluded due to miscarriage, stillbirth, premature birth, or low birth weight, leaving 319 participants for analysis (161 in the intervention group and 158 in the control group). A total of 32, 15, and 8 participants could not be contacted by phone for unknown reasons at the first, second, and third follow-ups, respectively. Moreover, 94 mothers missed the follow-up deadline (children <180 days age) at the 4-5 months postpartum follow-up because of the long holiday of Chinese Lunar New Year and the coronavirus (COVID-19) outbreak in January 2020 in China. The CONSORT (Consolidated Standards of Reporting Trials) [31] flow diagram is shown in Figure 2. Most demographic characteristics were similar between the participants who dropped out and participants who were followed up (Multimedia Appendix 1). However, at the 0-1 month postpartum follow-up, more mothers who only attended primary school dropped out; at the 4-5 months postpartum follow-up, more mothers who had a higher gestational age at enrollment dropped out.

Figure 2. CONSORT flow diagram. CONSORT: Consolidated Standards of Reporting Trials.



Characteristics of the Participants

The median age of enrolled participants was 28 years (IQR 25-31). There were no differences between the intervention and control groups for most demographic characteristics, except for gravidity and parity (Table 1). Compared with those in the control group, more participants in the intervention group had their first pregnancy (intervention: 46/161, 28.6%; control: 26/158, 16.5%; *P*=.01) or primipara (intervention: 48/161, 29.8%; control: 28/158, 17.7%; *P*=.01). The prevalence of maternal anemia was lower in the intervention group (64/161, 40.0%) than that in the control group (72/158, 45.6%), but the difference was not statistically significant (*P*=.32).

Table 1. Baseline characteristics of participants by treatment groups.

Characteristics	Total (N=319), n (%)	WeChat group (n=161), n (%)	Control group (n=158), n (%)	*P* value
Age (years), median (IQR)	28 (25-31)	28 (24-31)	28 (25-31)	.45
Gestational age (weeks)				**.27**
11-27	186 (58.3)	89 (55.3)	97 (61.4)	
28-42	133 (41.7)	72 (44.7)	61 (38.6)	
Gravidity				**.01**
First pregnancy	72 (22.6)	46 (28.6)	26 (16.5)	
Second pregnancy or more	247 (77.4)	115 (71.4)	132 (83.5)	
Parity				**.01**
Primipara	76 (23.8)	48 (29.8)	28 (17.7)	
Multipara	243 (76.2)	113 (70.2)	130 (82.3)	
Education level				**.62**
Primary school or below	54 (16.9)	24 (14.9)	30 (19.0)	
Middle school	173 (54.2)	89 (55.3)	84 (53.2)	
High school or above	92 (28.9)	48 (29.8)	44 (27.8)	
Occupation				**.75**
Stay-at-home	278 (87.1)	138 (85.7)	140 (88.6)	
Self-employed	9 (2.8)	5 (3.1)	4 (2.5)	
Farmer	5 (1.6)	2 (1.3)	3 (1.9)	
Others	27 (8.5)	16 (9.9)	11 (7.0)	
Nationality				**.86**
Han	244 (76.5)	125 (77.6)	119 (75.3)	
Tu	49 (15.4)	23 (14.3)	26 (16.5)	
Others	13 (8.1)	13 (8.1)	13 (8.2)	
Ever received antenatal care	252 (79.0)	121 (75.2)	131 (82.9)	.12
Anemia	136 (42.8)	64 (40.0)	72 (45.6)	.32

WeChat Activity

A total of 108 messages were published in the *Ke Xue Wei Yang* module in the WeChat official account, which was read more than 8892 times. The top 5 read messages were (1) benefits of breastfeeding (393 times); (2) importance of the early initiation (284 times); (3) breastfeeding positions and latching-on (242 times); (4) what is early initiation (214 times); (5) WHO recommendation: children should be exclusively breastfed from birth to 6 months (214 times).

Breastfeeding Practice and Knowledge

As illustrated in Table 2, nearly all children were breastfed in both groups. The early initiation rate was low (intervention: 93/148, 62.8%; control: 101/139, 72.7%); however, the difference was not statistically significant (*P*=.08). In addition, approximately 40% of children were given prelacteal feeding in both groups, with 25.8% (74/287) given infant formula and 16.7% (48/287) given water.

Table 2. Comparison of infant feeding practices.

Indicators	Total[a]	Intervention group[a]	Control group[a]	*P* value	Odds ratio (95% CI)	Adjusted odds ratio (95% CI)[b]
First follow-up (0-1 month)	**287 (100)**	**148 (100)**	**139 (100)**			
Early initiation	194 (67.6)	93 (62.8)	101 (72.7)	.11	0.64 (0.39-1.05)	0.66 (0.40-1.09)
Ever breastfeeding	283 (98.6)	147 (99.3)	136 (97.8)	.30	3.24 (0.33-31.54)	3.42 (0.34-34.4)
Prelacteal feeding	**122 (42.5)**	**59 (39.9)**	**63 (45.3)**	**.28**	**0.80 (0.50-1.28)**	**0.77 (0.48-1.24)**
Giving formula	74 (25.8)	37 (25.0)	37 (26.6)	N/A[c]	N/A	N/A
Giving water	48 (16.7)	22 (14.9)	26 (18.7)	N/A	N/A	N/A
Exclusive breastfeeding	208 (72.5)	120 (81.1)	88 (63.3)	<.001	2.49 (1.45-4.25)	2.75 (1.58-4.78)
Predominant breastfeeding	218 (76.0)	124 (83.8)	94 (67.6)	<.001	2.47 (1.41-4.34)	2.77 (1.55-4.96)
Any breastfeeding	279 (97.2)	146 (98.6)	133 (95.7)	.18	3.29 (0.65-16.59)	3.09 (0.61-15.79)
Giving water during the past 24 hours	17 (5.9)	6 (4.1)	11 (7.9)	.17	0.49 (0.18-1.34)	0.49 (0.17-1.35)
Giving dairy or dairy products during the past 24 hours	55 (19.2)	20 (13.5)	35 (25.2)	.005	0.46 (0.25-0.85)	0.40 (0.21-0.75)
Giving semisolid or solid foods during the past 24 hours	2 (0.7)	2 (1.4)	0	.17	N/A	N/A
Second follow-up (2-3 months)	**304 (100)**	**152 (100)**	**152 (100)**			
Exclusive breastfeeding	207 (68.1)	111 (73.0)	96 (63.2)	.09	1.47 (0.91, 2.38)	1.53 (0.94, 2.49)
Predominant breastfeeding	212 (69.7)	113 (74.3)	99 (65.1)	.07	1.55 (0.95, 2.54)	1.60 (0.96, 2.64)
Any breastfeeding	287 (94.4)	144 (94.7)	143 (94.1)	.64	1.13 (0.43-3.02)	1.27 (0.47-3.46)
Giving water during the past 24 hours	11 (3.6)	3 (2.0)	8 (5.3)	.18	0.36 (0.09-1.39)	0.40 (0.10-1.55)
Giving dairy or dairy products during the past 24 hours	66 (21.7)	28 (18.4)	38 (25.0)	.14	0.68 (0.39-1.18)	0.66 (0.38-1.15)
Giving semisolid or solid foods during the past 24 hours	5 (1.6)	2 (1.3)	3 (2.0)	.54	0.66 (0.10-4.02)	0.56 (0.09-3.56)
Third follow-up (4-5 months)	**217 (100)**	**108 (100)**	**109 (100)**			
Exclusive breastfeeding	96 (44.2)	50 (46.3)	46 (42.2)	.27	1.18 (0.69-2.02)	1.37 (0.78-2.39)
Predominant breastfeeding	111 (51.2)	58 (53.7)	53 (48.6)	.20	1.23 (0.72-2.09)	1.46 (0.82-2.51)
Any breastfeeding	198 (91.2)	101 (93.5)	97 (89.0)	.10	1.78 (0.68-4.72)	2.34 (0.84-6.54)
Giving water during the past 24 hours	27 (12.4)	12 (11.1)	15 (13.8)	.63	0.78 (0.34-1.76)	0.82 (0.36-1.87)
Giving dairy or dairy products during the past 24 hours	62 (28.6)	28 (25.9)	34 (31.2)	.12	0.77 (0.43-1.39)	0.61 (0.32-1.14)
Giving semisolid or solid foods during the past 24 hours	49 (22.6)	25 (23.2)	24 (22.0)	.94	1.07 (0.56-2.02)	1.02 (0.53-1.96)

[a] The number of participants varied because of loss to follow-up.

[b] Multiple logistic regression controlled for baseline parity.

[c] N/A: not applicable.

There was a downward trend for exclusive breastfeeding, predominant feeding, and any breastfeeding across the 3 follow-ups. At the 0-1 month postpartum follow-up, the exclusive breastfeeding rate was significantly higher in the intervention group (120/148, 81.1%) than in the control group (88/139, 63.3%), with an odds ratio (OR) of 2.75 (95% CI 1.58-4.78; *P*<.001). Similarly, mothers in the intervention group were more likely to provide breast milk predominantly (intervention: 124/148, 83.8%; control: 94/139, 67.6%; OR 2.77, 95% CI 1.55-4.96; *P*<.001), and less likely to give dairy or dairy products to their children (intervention: 20/148, 13.5%; control: 35/139, 25.2%; OR 0.40, 95% CI 0.21-0.75; *P*=.005).

At the 2-3 months postpartum follow-up, exclusive breastfeeding (intervention: 111/152, 73.0%; control: 96/152, 63.2%) and predominant feeding (intervention: 113/152, 74.3%; control: 99/152, 65.1%) in the intervention group were higher than in the control group, and fewer children were given dairy or dairy products in the intervention group (intervention: 28/152, 18.4%; control: 38/152, 25.0%). However, the differences between the 2 groups were not statistically significant (*P*=.09 for exclusive breastfeeding, *P*=.07 for predominant feeding, and *P*=.14 for giving dairy or dairy products).

At 4-5 months postpartum, both the rates of exclusive breastfeeding and predominant feeding dropped to 44.2% (96/217) and 51.2% (111/217), respectively, and the proportion of children who were given dairy or dairy products increased to 28.6% (62/217) for both groups; there were no significant differences between the 2 groups (*P*=.27 for exclusive breastfeeding, *P*=.20 for predominant feeding, and *P*=.12 for giving dairy or dairy products).

At the first 2 follow-ups, the rate of any breastfeeding was quite high, whereas very few children were given water and semisolid or solid foods. However, any breastfeeding dropped to around 90% and the proportion of children who were given water and semisolid or solid foods increased to 12.4% (27/217) and 22.6% (49/217), respectively, 4-5 months postpartum. There were no significant differences between the 2 groups (0-1 month postpartum: *P*=.18 for any breastfeeding rate, *P*=.17 for giving water, and *P*=0.17 for giving semisolid or solid foods; 2-3 months postpartum: *P*=.64 for any breastfeeding rate, *P*=.18 for giving water, and *P*=.54 for giving semisolid or solid foods; 4-5 months postpartum: *P*=.10 for any breastfeeding rate, *P*=.63 for giving water, and *P*=.94 for giving semisolid or solid foods).

Table 3 shows that all feeding knowledge indicators were low at baseline, with 33.9% (108/319) knowing early initiation of breastfeeding, 28.8% (92/319) knowing the duration of exclusive breastfeeding, 1.3% (4/319) knowing continued breastfeeding until 2 years of age, and 61.8% (197/319) knowing introduction of complementary foods at 6-8 months in both groups. Mothers' feeding knowledge was greatly improved at follow-ups (*P*<.001). However, there were no differences between the intervention group and the control group in baseline and each follow-up. The proportion of pregnant women who ever received breastfeeding information during pregnancy in both groups increased dramatically from only 25.7% (82/319) at baseline to more than 80%-90% of mothers who ever received breastfeeding information during pregnancy or after delivery (*P*<.001), however, there was also no difference between the 2 groups (0-1 month postpartum: *P*=.24, 2-3 months postpartum: *P*=.63, 4-5 months postpartum: *P*=.09). In addition, 16.3% (52/319) of pregnant women reported having ever received infant formula information during their pregnancy.

Table 3. Comparison of infant feeding knowledge between the intervention and control groups.

Outcomes	Groups			*P* value		
	All	Intervention	Control	Intervention vs control	Intervention, baseline vs follow-up	Control, baseline vs follow-up
Baseline	**319 (100)**	**161 (100)**	**158 (100)**			
Knowing early initiation of breastfeeding	108 (33.9)	58 (36.0)	50 (31.6)	.41	N/A[a]	N/A
Knowing the duration of exclusive breastfeeding	92 (28.8)	47 (29.2)	45 (28.5)	.89	N/A	N/A
Knowing continued breastfeeding until 2 years	4 (1.3)	2 (1.2)	2 (1.3)	.98	N/A	N/A
Knowing introduction of complementary foods at 6-8 months	197 (61.8)	101 (62.7)	96 (60.8)	.72	N/A	N/A
Women ever received breastfeeding information during pregnancy	82 (25.7)	41 (25.5)	41 (25.9)	.92	N/A	N/A
Women ever received infant formula information during pregnancy	52 (16.3)	29 (18.0)	23 (14.6)	.40	N/A	N/A
First follow-up (0-1 month)	**287 (100)**	**148 (100)**	**139 (100)**			
Knowing early initiation of breastfeeding	196 (68.3)	101 (68.2)	95 (68.4)	.99	<.001	<.001
Knowing the duration of exclusive breastfeeding	199 (69.3)	107 (72.3)	92 (66.2)	.26	<.001	<.001
Knowing continued breastfeeding until 2 years	47 (16.4)	24 (16.2)	23 (16.6)	.94	<.001	<.001
Knowing introduction of complementary foods at 6-8 months	240 (83.6)	123 (83.1)	117 (84.2)	.81	<.001	<.001
Women ever received breastfeeding information during pregnancy or after delivery	239 (83.3)	127 (85.8)	112 (80.6)	.24	<.001	<.001
Women ever received infant formula information during pregnancy or after delivery	14 (4.9)	9 (6.1)	5 (3.6)	.33	.001	.001
Second follow-up (2-3 months)	**304 (100)**	**152 (100)**	**152 (100)**			
Knowing early initiation of breastfeeding	216 (71.1)	115 (75.7)	101 (66.5)	.08	<.001	<.001
Knowing the duration of exclusive breastfeeding	240 (78.9)	122 (80.3)	118 (77.6)	.57	<.001	<.001
Knowing continued breastfeeding until two years	86 (28.3)	44 (28.9)	42 (27.6)	.80	<.001	<.001
Knowing introduction of complementary foods at 6-8 months	262 (86.2)	136 (89.5)	126 (82.9)	.10	<.001	<.001
Women ever received breastfeeding information during pregnancy or after delivery	286 (94.1)	144 (94.7)	142 (93.4)	.63	<.001	<.001
Women ever received infant formula information during pregnancy or after delivery	10 (3.3)	3 (2.0)	7 (4.6)	.20	<.001	.003
Third follow-up (4-5 months)	**217 (100)**	**108 (100)**	**109 (100)**			
Knowing early initiation of breastfeeding	171 (78.8)	84 (77.8)	87 (79.8)	.71	<.001	<.001
Knowing the duration of exclusive breastfeeding	194 (89.4)	94 (87.0)	100 (91.7)	.26	<.001	<.001
Knowing continued breastfeeding until 2 years	80 (36.9)	44 (40.7)	36 (33.0)	.24	<.001	<.001
Knowing introduction of complementary foods at 6-8 months	205 (94.5)	102 (94.4)	103 (94.5)	.99	<.001	<.001
Women ever received breastfeeding information during pregnancy or after delivery	208 (95.9)	101 (93.5)	107 (98.2)	.09	<.001	<.001
Women ever received infant formula information during pregnancy or after delivery	7 (3.2)	3 (2.8)	4 (3.7)	.71	<.001	.004

[a] N/A: not applicable.

Discussion

Principal Findings

This is the first randomized controlled trial using WeChat to promote exclusive breastfeeding in rural China. This study showed that antenatal plus postnatal WeChat breastfeeding education was associated with higher rates of exclusive and predominant breastfeeding in the early postnatal period. Mothers in the WeChat group had 2.7 times the odds of exclusive and predominant breastfeeding during the first 2 months postpartum. We found that giving dairy or dairy products, water, and semisolid or solid foods were common reasons for nonexclusive breastfeeding and that giving dairy or dairy products was the predominant reason, with 19.2% (55/287), 21.7% (66/304), and 28.6% (62/217) of all children being given dairy or dairy products (0-1 months, 2-3 months, and 4-5 months postpartum, respectively). In addition, 25.8% of newborns (74/287) were given prelacteal infant formula, which can limit an infant's frequency of suckling and expose them to increased risk of infection [32]. Our study demonstrated that a WeChat intervention could significantly reduce dairy or dairy product supplementation in early life, and thus improve the exclusive breastfeeding rate. The proportion in the WeChat group giving dairy or dairy products 0-1 month postpartum was significantly lower than those in the control group (*P*=.005). Although there was no significant difference between the groups 2-3 months (*P*=.14) or 4-5 months postpartum (*P*=.12), fewer children in the WeChat group were given dairy or dairy products.

Proportions of children who were given water were 5.9% (17/287), 3.6% (11/304), and 12.4% (27/217) (0-1 months, 2-3 months, and 4-5 months postpartum, respectively), and no difference was found between the groups. Very few children were given semisolid or solid foods in the first 2 follow-ups; however, the proportion dramatically increased to 22.6% (49/217) 4-5 months postpartum. Similarly, for children being given water, we found no difference between groups at all 3 follow-ups. Therefore, the WeChat intervention in our study could significantly reduce dairy or dairy product supplementation.

Comparison With Prior Work

It has been demonstrated that information and communication systems, such as web platform, mobile apps, and SMS can be used to promote, educate, and support breastfeeding women, and offer effective means of improving breastfeeding outcomes [33,34]. From 2010 to 2012, a study in Shanghai using SMS to promote breastfeeding found that weekly messages on infant feeding from the third trimester to 12 months postpartum could improve the exclusive breastfeeding rate at 6 months (OR 2.67) but not at 4 months [35]. This explained that many mothers in China were unaware the new concept of exclusive breastfeeding to 6 months of age, and the SMS disseminated the information to mothers [35]. In contrast, the WeChat intervention in our study improved the exclusive breastfeeding rate during the first 2 months postpartum and mainly reduced infant formula instead of water and semisolid or solid food. One reason might be that, in the first 2 months of breastfeeding, mothers may encounter more breastfeeding problems such as perceived insufficient milk supply, breast engorgement, poor suckling technique (of the infant), and sore nipples, which are well-known predictors for early formula supplementation and breastfeeding cessation [36-38]. Therefore, sufficient professional breastfeeding information is needed to increase their breastfeeding confidence during this period. However, 2 months postpartum, mothers may face substantial pressure from family and social culture to introduce formula or to wean [39], which may explain the limited effect on breastfeeding behavior changes afterward.

Improvements in health outcomes rely on putting knowledge into practice; however, population studies have documented that there is a gap between expectations and the actual performance of behaviors in health care and prevention [40]. Compared with the baseline, mothers' feeding knowledge improved at the 3 follow-ups for both intervention and control groups in this study. Proportions of mothers knowing the duration of exclusive breastfeeding dramatically increased from 28.8% (92/319) at baseline to 69.3% (199/287), 78.9% (240/304), and 89.4% (194/217) (0-1 months, 2-3 months, and 4-5 months postpartum, respectively). In contrast, the exclusive breastfeeding rate dropped from 72.5% (208/287, 0-1 month postpartum) to 44.2% (96/217, 4-5 months postpartum). This is in line with the findings of a previous study [41] that described implementing their breastfeeding promotion interventions in the face of "time and space burdens." In addition, no differences in mothers' knowledge were found between the 2 groups, as expected, throughout the study. The reason might be that although mothers in the control group had no access to our WeChat information, they still had various sources of feeding information, including health facilities, mass media, books, and the internet. Successful breastfeeding promotion needs not only education but also support [9,10]. Therefore, multichannel support, such as experts online, through mobilized community, and by health facilities, should also be given to mothers.

With the widespread use of WeChat, it has become a potential health promotion tool in China [22-27,42-44]. The WeChat intervention in our study, providing mothers with information on breastfeeding promotion, could improve breastfeeding practices. The results were in accordance with those of previous studies on weight loss [27] and malaria [23], both of which showed that participants' knowledge, attitudes, and practices were greatly changed via WeChat official accounts. Information on the benefits of breastfeeding, early initiation, breastfeeding positions, and latching-on were read most during the intervention period, which indicated that these themes were primarily concerned among mothers. However, the effect is unsustainable for a longer time. Therefore, future studies should focus on how to improve the sustainability of the effect.

Limitations

Our study had several limitations. First, as the Chinese Lunar Year and the COVID-19 outbreak occurred in January 2020 in China, interviewers could not conduct follow-up data collection during that time, which led to a total of 94 mother-child pairs missing the 4-5 months postpartum follow-up deadline (180 days). However, we compared the baseline demographic characteristics between the participants who dropped out and those who were followed up and did not find differences between 2 groups, except for gestational age. Second, given the popular use of WeChat, there are other WeChat official accounts on breastfeeding promotion, and we cannot guarantee that each participant in both groups did not receive breastfeeding knowledge from other sources during the intervention period. However, our randomized controlled trial design could limit such bias. Third, as the randomization unit was individual pregnant women, contamination may exist between intervention and control groups within the same township.

Conclusions

This health education intervention for promoting breastfeeding via WeChat official accounts was associated with the improvement of exclusive breastfeeding rate in early life. This health education intervention strategy can be used as a reference for local breastfeeding promotion programs, especially, in rural areas of western China where fewer high-quality health services are provided than in urban areas.

Acknowledgments

The study was funded by UNICEF. The authors wish to thank all colleagues from the Huzhu Maternal and Child Health and Family Planning Service Center for coordination, logistic arrangements, and conducting follow up interviews. We would also like to thank all students from the Qinghai Institute of Health Sciences for their hard work as baseline interviewers. In addition, we would like to thank ZYZY (Beijing) Pioneer of Cultural Essence Co Ltd for providing their expertise developing the enhanced WeChat platform. Finally, we are indebted to all the mothers who participated in our study.

Authors' Contributions

YZ, QW, MV, WW, and YH designed the study; YZ, QW, and YH developed the intervention and contributed to implementing the study; YH conducted data collection; QW, YH, ZL, and WW analyzed and interpreted data. QW wrote the first draft of the manuscript, which was substantially revised by YZ, WW, MV, and ZL; all authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of characteristics between participants with complete and incomplete data.
[DOCX File , 18 KB - jmir_v22i12e23273_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 11107 KB - jmir_v22i12e23273_app2.pdf]

References

1. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, Maternal Child Nutrition Study Group. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet 2013 Aug 03;382(9890):427-451. [doi: 10.1016/S0140-6736(13)60937-X] [Medline: 23746772]
2. Dewey KG, Adu-Afarwuah S. Systematic review of the efficacy and effectiveness of complementary feeding interventions in developing countries. Matern Child Nutr 2008 Apr;4 Suppl 1:24-85 [FREE Full text] [doi: 10.1111/j.1740-8709.2007.00124.x] [Medline: 18289157]
3. Stein AD, Thompson AM, Waters A. Childhood growth and chronic disease: evidence from countries undergoing the nutrition transition. Matern Child Nutr 2005 Jul;1(3):177-184 [FREE Full text] [doi: 10.1111/j.1740-8709.2005.00021.x] [Medline: 16881898]
4. Global Targets 2025. World Health Organization. URL: https://www.who.int/nutrition/global-target-2025/en/ [accessed 2020-04-10]
5. Global strategy for women's and children's health 2010. World Health Organization. URL: https://www.who.int/pmnch/topics/maternal/20100914_gswch_en.pdf [accessed 2020-04-10]
6. Center for Health Statistics and Information, Ministry of Health China. An Analysis Report of National Health Services Survey in China. Beijing: Peking Union Medical College Press; Dec 2009.
7. Duan Y, Yang Z, Lai J, Yu D, Chang S, Pang X, et al. Exclusive breastfeeding rate and complementary feeding indicators in China: a national representative survey in 2013. Nutrients 2018 Feb 22;10(2):249 [FREE Full text] [doi: 10.3390/nu10020249] [Medline: 29470415]
8. Spiby H, McCormick F, Wallace L, Renfrew MJ, D'Souza L, Dyson L. A systematic review of education and evidence-based practice interventions with health professionals and breast feeding counsellors on duration of breast feeding. Midwifery 2009 Feb;25(1):50-61. [doi: 10.1016/j.midw.2007.01.006] [Medline: 17418464]

9. Skouteris H, Bailey C, Nagle C, Hauck Y, Bruce L, Morris H. Interventions designed to promote exclusive breastfeeding in high-income countries: a systematic review update. Breastfeed Med 2017 Dec;12(10):604-614. [doi: 10.1089/bfm.2017.0065] [Medline: 28885859]
10. Skouteris H, Nagle C, Fowler M, Kent B, Sahota P, Morris H. Interventions designed to promote exclusive breastfeeding in high-income countries: a systematic review. Breastfeed Med 2014 Apr;9(3):113-127. [doi: 10.1089/bfm.2013.0081] [Medline: 24568270]
11. Implementation guidelines on national basic public health service program. Ministry of Health China. 2009 Oct 16. URL: http://www.nhc.gov.cn/jws/s3581r/200910/fe1cdd87dcfa4622abca696c712d77e8.shtml [accessed 2020-11-26]
12. Wu Q, Scherpbier RW, van Velthoven MH, Chen L, Wang W, Li Y, et al. Poor infant and young child feeding practices and sources of caregivers' feeding knowledge in rural Hebei Province, China: findings from a cross-sectional survey. BMJ Open 2014 Jul 29;4(7):e005108 [FREE Full text] [doi: 10.1136/bmjopen-2014-005108] [Medline: 25079931]
13. Yue A, Marsh L, Zhou H, Medina A, Luo R, Shi Y, et al. Nutritional deficiencies, the absence of information and caregiver shortcomings: a qualitative analysis of infant feeding practices in rural China. PLoS One 2016;11(4):e0153385 [FREE Full text] [doi: 10.1371/journal.pone.0153385] [Medline: 27073864]
14. Yue A, Zhang N, Liu X, Tang L, Luo R, Yang M, et al. Do infant feeding practices differ between grandmothers and mothers in rural China? evidence from rural Shaanxi Province. Fam Community Health 2018;41(4):233-243. [doi: 10.1097/FCH.0000000000000198] [Medline: 30134338]
15. Fralick M, Haj R, Hirpara D, Wong K, Muller M, Matukas L, et al. Can a smartphone app improve medical trainees' knowledge of antibiotics? Int J Med Educ 2017 Nov 30;8:416-420 [FREE Full text] [doi: 10.5116/ijme.5a11.8422] [Medline: 29200402]
16. Plotnikoff RC, Wilczynska M, Cohen KE, Smith JJ, Lubans DR. Integrating smartphone technology, social support and the outdoor physical environment to improve fitness among adults at risk of, or diagnosed with, type 2 diabetes: findings from the 'eCoFit' randomized controlled trial. Prev Med 2017 Dec;105:404-411. [doi: 10.1016/j.ypmed.2017.08.027] [Medline: 28887192]
17. Gajecki M, Andersson C, Rosendahl I, Sinadinovic K, Fredriksson M, Berman AH. Skills training via smartphone app for university students with excessive alcohol consumption: a randomized controlled trial. Int J Behav Med 2017 Oct;24(5):778-788 [FREE Full text] [doi: 10.1007/s12529-016-9629-9] [Medline: 28224445]
18. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
19. Schoeppe S, Alley S, Van LW, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
20. 6 Powerful WeChat Statistics in 2020 You Need to Know. Brew Interactive. 2020. URL: https://blog.brewinteractive.com/wechat-statistics [accessed 2020-05-17]
21. Feiran W. WeChat official accounts –an “app-within-an-app” platform. Digital Initiative. URL: https://digital.hbs.edu/platform-rctom/submission/wechat-official-accounts-an-app-within-an-app-platform/ [accessed 2020-05-17]
22. Zhang X, Xiao H, Chen Y. Evaluation of a WeChat-based life review programme for cancer patients: A quasi-experimental study. J Adv Nurs 2019 Jul;75(7):1563-1574. [doi: 10.1111/jan.14018] [Medline: 30950535]
23. Li W, Han LQ, Guo YJ, Sun J. Using WeChat official accounts to improve malaria health literacy among Chinese expatriates in Niger: an intervention study. Malar J 2016 Nov 24;15(1):567 [FREE Full text] [doi: 10.1186/s12936-016-1621-y] [Medline: 27881122]
24. Cao Y, Lin S, Zhu D, Xu F, Chen Z, Shen H, et al. WeChat public account use improves clinical control of cough-variant asthma: a randomized controlled trial. Med Sci Monit 2018 Mar 14;24:1524-1532 [FREE Full text] [doi: 10.12659/msm.907284] [Medline: 29536984]
25. Feng S, Liang Z, Zhang R, Liao W, Chen Y, Fan Y, et al. Effects of mobile phone WeChat services improve adherence to corticosteroid nasal spray treatment for chronic rhinosinusitis after functional endoscopic sinus surgery: a 3-month follow-up study. Eur Arch Otorhinolaryngol 2017 Mar;274(3):1477-1485. [doi: 10.1007/s00405-016-4371-0] [Medline: 27822702]
26. Dong Y, Wang P, Dai Z, Liu K, Jin Y, Li A, et al. Increased self-care activities and glycemic control rate in relation to health education via Wechat among diabetes patients: a randomized clinical trial. Medicine (Baltimore) 2018 Dec;97(50):e13632 [FREE Full text] [doi: 10.1097/MD.0000000000013632] [Medline: 30558051]
27. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social media-promoted weight loss among an occupational population: cohort study using a WeChat mobile phone app-based campaign. J Med Internet Res 2017 Oct 23;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
28. Wu Q, Huang Y, van Velthoven MH, Wang W, Chang S, Zhang Y. The effectiveness of using a WeChat account to improve exclusive breastfeeding in Huzhu County Qinghai Province, China: protocol for a randomized control trial. BMC Public Health 2019 Dec 02;19(1):1603 [FREE Full text] [doi: 10.1186/s12889-019-7676-2] [Medline: 31791295]
29. General introduction of Huzhu County. Huzhu County Government. 2020. URL: http://www.huzhu.gov.cn/zmhz/hzgk.htm [accessed 2020-07-24]

30. WHO, UNICEF, USAID, AED, IFPRI. Indicators for assessing infant and young child feeding practices part I: definition. World Health Organization. 2010. URL: https://www.who.int/maternal_child_adolescent/documents/9789241596664/en/ [accessed 2020-05-20]
31. Eysenbach G, CONSORT- E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
32. Meshram II. Impact of feeding and breastfeeding practices on the nutritional status of infants in a district of Andhra Pradesh, India. Natl Med J India 2012;25(4):201-206. [Medline: 23278776]
33. Tang K, Gerling K, Chen W, Geurts L. Information and communication systems to tackle barriers to breastfeeding: systematic search and review. J Med Internet Res 2019 Sep 27;21(9):e13947 [FREE Full text] [doi: 10.2196/13947] [Medline: 31573903]
34. Almohanna AA, Win KT, Meedya S. Effectiveness of internet-based electronic technology interventions on breastfeeding outcomes: systematic review. J Med Internet Res 2020 May 29;22(5):e17361 [FREE Full text] [doi: 10.2196/17361] [Medline: 32469315]
35. Jiang H, Li M, Wen LM, Hu Q, Yang D, He G, et al. Effect of short message service on infant feeding practice: findings from a community-based study in Shanghai, China. JAMA Pediatr 2014 May;168(5):471-478 [FREE Full text] [doi: 10.1001/jamapediatrics.2014.58] [Medline: 24639004]
36. Hegney D, Fallon T, O'Brien ML. Against all odds: a retrospective case-controlled study of women who experienced extraordinary breastfeeding problems. J Clin Nurs 2008 May;17(9):1182-1192. [doi: 10.1111/j.1365-2702.2008.02300.x] [Medline: 18416794]
37. Parry JE, Ip DKM, Chau PYK, Wu KM, Tarrant M. Predictors and consequences of in-hospital formula supplementation for healthy breastfeeding newborns. J Hum Lact 2013 Nov;29(4):527-536. [doi: 10.1177/0890334412474719] [Medline: 23439865]
38. Beyene AM, Liben ML, Arora A. Factors associated with the early termination of exclusive breastfeeding among mother-infant dyads in Samara-Logia, Northeastern Ethiopia. BMC Pediatr 2019 Nov 11;19(1):428 [FREE Full text] [doi: 10.1186/s12887-019-1803-1] [Medline: 31711461]
39. Zhang K, Tang L, Wang H, Qiu L, Binns CW, Lee AH. Why do mothers of young infants choose to formula feed in China? Perceptions of mothers and hospital staff. Int J Environ Res Public Health 2015 Apr 24;12(5):4520-4532 [FREE Full text] [doi: 10.3390/ijerph120504520] [Medline: 25918908]
40. Adams RJ. Improving health outcomes with better patient understanding and education. Risk Manag Healthc Policy 2010;3:61-72 [FREE Full text] [doi: 10.2147/RMHP.S7500] [Medline: 22312219]
41. Bonuck KA, Lischewski J, Brittner M. Clinical translational research hits the road: RCT of breastfeeding promotion interventions in routine prenatal care. Contemp Clin Trials 2009 Sep;30(5):419-426 [FREE Full text] [doi: 10.1016/j.cct.2009.05.005] [Medline: 19523539]
42. Zhang Y, Fan D, Ji H, Qiao S, Li X. Correction: treatment adherence and secondary prevention of ischemic stroke among discharged patients using mobile phone- and WeChat-based improvement services: cohort study. JMIR Mhealth Uhealth 2020 Apr 29;8(4):e19454 [FREE Full text] [doi: 10.2196/19454] [Medline: 32348276]
43. Jiang Y, Liu F, Guo J, Sun P, Chen Z, Li J, et al. Evaluating an intervention program using WeChat for patients with chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2020 Apr 21;22(4):e17089 [FREE Full text] [doi: 10.2196/17089] [Medline: 32314971]
44. Guo Y, Hong YA, Cai W, Li L, Hao Y, Qiao J, et al. Effect of a WeChat-based intervention (Run4Love) on depressive symptoms among people living with HIV in China: a randomized controlled trial. J Med Internet Res 2020 Feb 09;22(2):e16715 [FREE Full text] [doi: 10.2196/16715] [Medline: 32044751]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
OR: odds ratio
SMS: short message service
UNICEF: United Nations Children's Fund
WHO: World Health Organization

Edited by G Eysenbach, R Kukafka; submitted 13.08.20; peer-reviewed by J Huo, C Calyx; comments to author 01.09.20; revised version received 06.09.20; accepted 02.11.20; published 03.12.20.

Please cite as:
Wu Q, Huang Y, Liao Z, van Velthoven MH, Wang W, Zhang Y
Effectiveness of WeChat for Improving Exclusive Breastfeeding in Huzhu County China: Randomized Controlled Trial
J Med Internet Res 2020;22(12):e23273
URL: https://www.jmir.org/2020/12/e23273
doi:10.2196/23273
PMID:33270026

Original Paper

A Remote Nutritional Intervention to Change the Dietary Habits of Patients Undergoing Ablation of Atrial Fibrillation: Randomized Controlled Trial

Leticia Goni[1,2], PhD; Víctor de la O[1,2], MSc; M Teresa Barrio-López[3], PhD, MD; Pablo Ramos[4], MD; Luis Tercedor[5,6], MD; Jose Luis Ibañez-Criado[7], PhD, MD; Eduardo Castellanos[3], PhD, MD; Alicia Ibañez Criado[7], MD; Rosa Macias Ruiz[5], MD; Ignacio García-Bolao[4], PhD, FESC, MD; Jesus Almendral[3], PhD, MD, FESC; Miguel Ángel Martínez-González[1,2,8], PhD, MD, MPH; Miguel Ruiz-Canela[1], MPH, PhD

[1]Department of Preventive Medicine and Public Health, University of Navarra, Instituto de Investigación Sanitaria de Navarra, Pamplona, Spain
[2]Fisiopatología de la Obesidad y Nutrición, Centro de Investigación Biomédica en Red, Instituto de Salud Carlos III, Madrid, Spain
[3]Electrophysiology Laboratory and Arrhythmia Unit, Hospital Montepríncipe, Grupo HM Hospitales, University CEU-San Pablo, Madrid, Spain
[4]Arrhythmia Unit, Department of Cardiology and Cardiac Surgery, Clínica Universidad de Navarra, Pamplona, Spain
[5]Department of Cardiology, Virgen de las Nieves University Hospital, Granada, Spain
[6]Biosanitary Research Institute of Granada (ibs.GRANADA), Granada, Spain
[7]Arrhythmia Unit, Cardiology Service, Alicante Institute of Health and Biomedical Research (ISABIAL-FISABIO Foundation), University General Hospital of Alicante, Alicante, Spain
[8]Department of Nutrition, Harvard TH Chan School of Public Health, Harvard University, Boston, MA, United States

Corresponding Author:
Miguel Ruiz-Canela, MPH, PhD
Department of Preventive Medicine and Public Health
University of Navarra
C/ Irunlarrea, 1
Pamplona
Spain
Phone: 34 948425600 ext 806395
Email: mcanela@unav.es

Abstract

Background: The Prevention With Mediterranean Diet (PREDIMED) trial supported the effectiveness of a nutritional intervention conducted by a dietitian to prevent cardiovascular disease. However, the effect of a remote intervention to follow the Mediterranean diet has been less explored.

Objective: This study aims to assess the effectiveness of a remotely provided Mediterranean diet–based nutritional intervention in obtaining favorable dietary changes in the context of a secondary prevention trial of atrial fibrillation (AF).

Methods: The PREvention of recurrent arrhythmias with Mediterranean diet (PREDIMAR) study is a 2-year multicenter, randomized, controlled, single-blinded trial to assess the effect of the Mediterranean diet enriched with extra virgin olive oil (EVOO) on the prevention of atrial tachyarrhythmia recurrence after catheter ablation. Participants in sinus rhythm after ablation were randomly assigned to an intervention group (Mediterranean diet enriched with EVOO) or a control group (usual clinical care). The remote nutritional intervention included phone contacts (1 per 3 months) and web-based interventions with provision of dietary recommendations, and participants had access to a web page, a mobile app, and printed resources. The information is divided into 6 areas: *Recommended foods*, *Menus*, *News and Online resources*, *Practical tips*, *Mediterranean diet classroom*, and *Your personal experience*. At baseline and at 1-year and 2-year follow-up, the 14-item Mediterranean Diet Adherence Screener (MEDAS) questionnaire and a semiquantitative food frequency questionnaire were collected by a dietitian by phone.

Results: A total of 720 subjects were randomized (365 to the intervention group, 355 to the control group). Up to September 2020, 560 subjects completed the first year (560/574, retention rate 95.6%) and 304 completed the second year (304/322, retention rate 94.4%) of the intervention. After 24 months of follow-up, increased adherence to the Mediterranean diet was observed in both groups, but the improvement was significantly higher in the intervention group than in the control group (net between-group difference: 1.8 points in the MEDAS questionnaire (95% CI 1.4-2.2; *P*<.001). Compared with the control group, the Mediterranean diet intervention group showed a significant increase in the consumption of fruits (*P*<.001), olive oil (*P*<.001), whole grain cereals (*P*=.002), pulses (*P*<.001), nuts (*P*<.001), white fish (*P*<.001), fatty fish (*P*<.001), and white meat (*P*=.007), and a significant reduction in refined cereals (*P*<.001), red and processed meat (*P*<.001), and sweets (*P*<.001) at 2 years of intervention. In terms of nutrients, the intervention group significantly increased their intake of omega-3 (*P*<.001) and fiber (*P*<.001), and they decreased their intake of carbohydrates (*P*=.02) and saturated fatty acids (*P*<.001) compared with the control group.

Conclusions: The remote nutritional intervention using a website and phone calls seems to be effective in increasing adherence to the Mediterranean diet pattern among AF patients treated with catheter ablation.

Trial Registration: ClinicalTrials.gov NCT03053843; https://www.clinicaltrials.gov/ct2/show/NCT03053843

(J Med Internet Res 2020;22(12):e21436) doi:10.2196/21436

KEYWORDS

atrial fibrillation; secondary prevention; remote intervention; Mediterranean diet; olive oil

Introduction

Atrial Fibrillation

Atrial fibrillation (AF) is currently the most common cardiac arrhythmia. In 2010, AF affected more than 33.5 million persons worldwide (20.9 million men and 12.6 million women) [1] and by 2030, 14 to 17 million people in Europe, and 12.1 million people in the United States will be diagnosed with AF [2,3]. This increased number of incident cases and age-adjusted prevalence of AF over the last few decades has led to a substantial disease and economic burden [1,4]. Currently, there are effective rhythm control strategies, such as catheter ablation, to recover sinus rhythm in patients with AF [5]. However, relapses occur in approximately 30% of the patients 1 year after the intervention [6]. A recent study showed that about 1 in 8 patients treated with catheter ablation needs to undergo a second procedure [7]. Therefore, research is needed to identify modifiable risk factors for the recurrence of AF to prevent recurrence and to maximize the durability of the sinus rhythm after ablation [8].

The Mediterranean Diet and AF

The Mediterranean diet is considered an ideal nutritional model for cardiovascular health [9]. However, data on the relationship between the Mediterranean diet and AF are limited. The Prevention With Mediterranean Diet (PREDIMED) trial, in primary prevention, showed a protective effect of a Mediterranean diet enriched with extra virgin olive oil (EVOO) on new-onset AF [10]. Previously, a short-term trial showed that a supplement of omega-3 fatty acids and vitamin antioxidants was associated with a lower probability of AF occurring after on-pump cardiac surgery [11]. A case-control study has also suggested that the Mediterranean diet and a high intake of antioxidants increased the probability of spontaneous conversion of AF [12]. Other cohort studies have analyzed the effect of some components of the Mediterranean diet, such as olive oil, red wine, fatty fish, and nuts, although with inconclusive results [13-16]. A higher consumption of coffee and chocolate is inversely associated with the risk of incident AF [17,18]. In summary, the anti-inflammatory and anti-oxidative effects of these foods seem to support the potential preventive effect of the Mediterranean diet on the development of AF [19].

As far as we know, no previous study has assessed the effect of a Mediterranean diet intervention on preventing recurrences of AF after ablation. The PREvention of recurrent arrhythmias with Mediterranean diet (PREDIMAR) study is an ongoing secondary prevention trial aimed at assessing the effect of a Mediterranean diet enriched with EVOO [20]. The novelty of this trial is that dietitians conduct the nutritional intervention remotely in coordination with the face-to-face intervention conducted by the cardiologists and health care team. Different studies have demonstrated that remote nutritional interventions (web page, mobile phone app, email, text messaging, and phone calls) are at least as effective as face-to-face interventions in terms of weight loss and in changing eating behaviors [21-23]. However, no previous study has assessed how a remote intervention may change the adherence to the Mediterranean diet in patients with AF after undergoing catheter ablation to recover sinus rhythm.

The principal objective of this study is to assess the effect of a web-based and telephone intervention in obtaining favorable dietary changes in the context of a secondary prevention trial of AF. In addition, we provide a detailed description of the remote nutritional intervention conducted in the PREDIMAR trial.

Methods

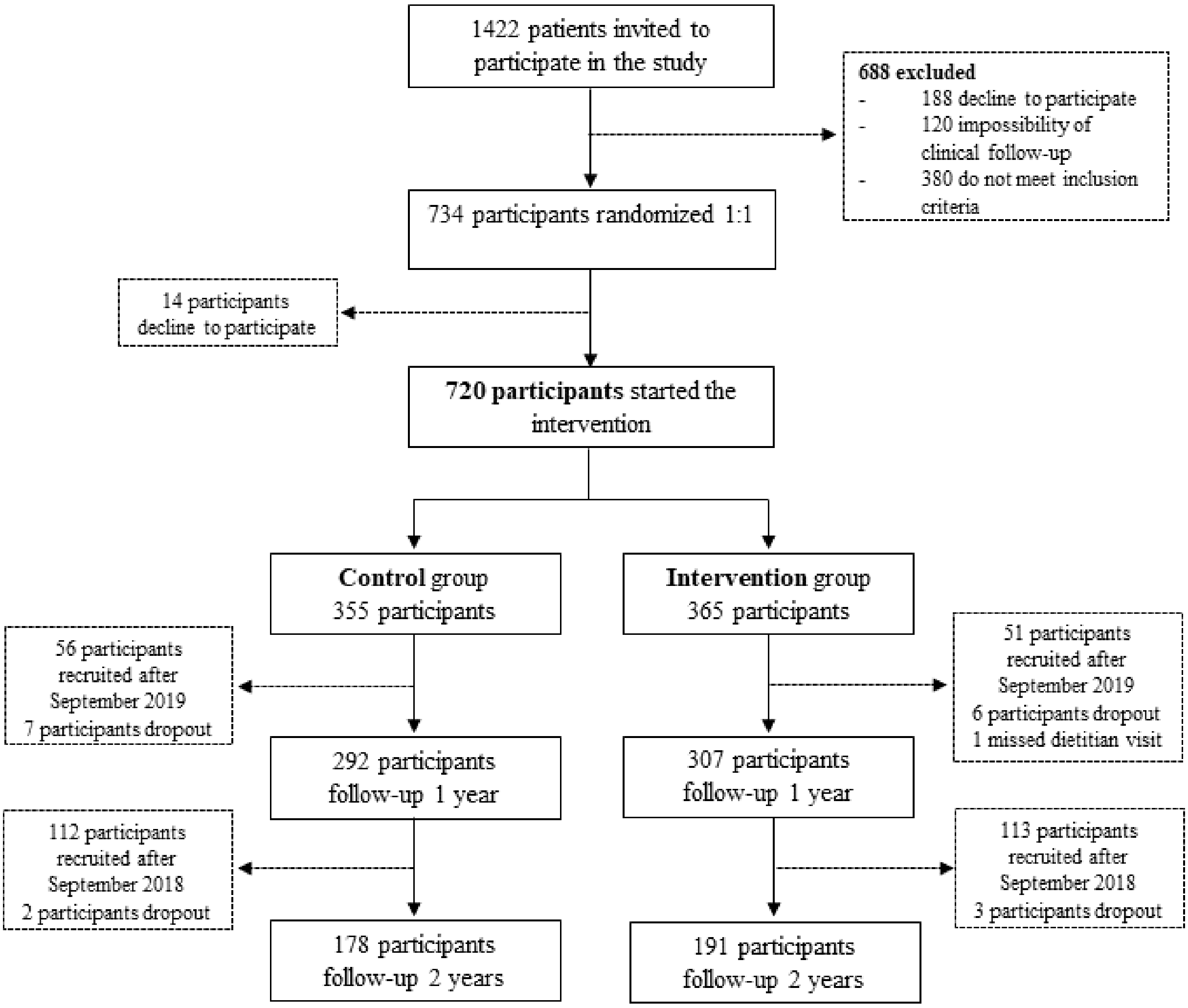
Overview of the PREDIMAR Study

The PREDIMAR study was a multicenter, randomized, controlled, single-blind trial. The study protocol has been described in detail elsewhere [20]. Briefly, the aim of the PREDIMAR study was to analyze the effect of a Mediterranean diet intervention enriched with EVOO on the prevention of atrial tachyarrhythmia recurrence after catheter ablation. Participants were recruited from 4 Spanish hospitals: Hospital Montepríncipe (Madrid), Clínica Universidad de Navarra (Pamplona), Hospital Virgen de las Nieves (Granada), and Hospital General Universitario (Alicante).

Between March 2017 and January 2020, 1422 patients were invited to participate in the study (Figure 1). Of them, although 734 patients were recruited and randomized, 720 started the intervention. The PREDIMAR trial is an ongoing study that will finish in January 2022. The flowchart shows participants who have completed 1- or 2-year follow-up with data updated to September 2020. Until this date, the number of dropouts was

18, and the retention rate was 95.6% among participants with follow-up over 12 months (560/574), and 94.4% with follow-up over 24 months (304/322).

Figure 1. Flowchart of participant screening, recruitment, and randomization.



After catheter ablation for the treatment of AF, all participants who gave their informed consent, and who fulfilled the inclusion criteria were randomly assigned to the intervention (Mediterranean diet enriched with EVOO) or the control group (usual clinical care). The intervention period took 2 years (Multimedia Appendix 1). During this time, clinical follow-up visits took place at 3, 6, 12, 18, and 24 months after the ablation. To determine possible atrial tachyarrhythmia recurrences, each patient received a portable cardiac rhythm-monitoring device on the third month visit, which was returned on the 18th-month visit. This was because recurrences that occur during the first 3 months after the ablation were not considered clinically relevant, as they could be due to the ablation and the healing process (so called blanking period) [5-7]. At baseline and at 12- and 24-month follow-up, information about lifestyle habits (including food habits, physical activity, and life quality) was recorded by a dietitian by phone. Blood and urine samples were also collected at baseline and at 12 and 24 months of the intervention period. Clinical care providers (cardiac electrophysiologists) were blinded to the group assignment.

The trial was registered at ClinicalTrials.gov NCT03053843. The Research Ethics Committees from each recruitment center approved the protocol. All participants provided written informed consent after they received the information sheet and additional verbal explanation of the study characteristics.

Nutritional Intervention

The aim of the nutritional intervention of the PREDIMAR study is to improve adherence to the Mediterranean diet. The Mediterranean diet is characterized by the exclusive use of EVOO for all culinary purposes and high consumption of vegetables, fruits, whole grains, legumes, and nuts; moderate consumption of fish; and very low consumption of red and processed meats, refined grains, sweet desserts, and whole-fat dairy products (only fermented dairy products, cheese and yogurt, are consumed in moderation) and ultraprocessed foods [9]. Specifically, the dietary recommendations for the intervention group were the use of 4 or more spoons of EVOO for cooking and dressing of dishes; consumption of 2 or more servings (200 g per serving) per day of vegetables (at least one of them as salads); 3 or more servings (125 g per serving) per day of fruits (including natural juices); 3 or more servings (60-80 g per serving) per week of legumes; 3 or more servings (150 g per serving) per week of fish or seafood (at least one serving of fatty fish); 3 or more servings (30 g per serving) per week of nuts; selected white meats (ie, poultry without skin, rabbit) instead of red meats (ie, beef, pork) or processed meats (ie, sausages, burgers); regularly cooking (2 or more times per week) with salsa made with minced tomato, garlic, and onion simmered in olive oil (sofrito) for dressing different dishes; selected whole grain cereals (ie, bread, pasta, rice) instead of refined cereals; eliminate or limit the consumption of cream, butter, and margarine, carbonated and/or sweetened beverages, commercial bakery products (ie, sweet desserts, cakes, pastries, cookies), and ultraprocessed foods. These recommendations were discussed with participants during periodical phone interviews with a dietitian, and personalized goals were set for the participants according to their improvement needs. Nutritional intervention was not based on a specific amount of calories or a macronutrient distribution.

In the PREDIMAR study, we used a remote nutritional intervention rather than an in-person approach. This remote intervention was conducted by a team of registered dietitians by phone and on the internet, and participants had access to web-based information on a website, a mobile app, and printed material. The intervention was conducted by the same group of dietitians to all participants from the 4 centers. This characteristic allowed a highly homogeneous intervention according to the protocol of the study, but it also allowed for tailored nutrition education through personal interviews conducted by phone and web-based communication with the dietitians.

Web Page

A multidisciplinary group of dietitians-nutritionists, epidemiologists, medical doctors, and chefs developed the content of the website for this study, and professional multimedia programmers produced it (Nubba Group).

Once a participant was randomized, they received an automated email with a username and password to access the PREDIMAR website. Participants in the control group had access to general information about AF only.

For participants in the intervention group, the content of the website was divided into 6 areas: *recommended foods*, *menus*, *News and Online resources*, *practical tips*, *Mediterranean Diet classroom*, and *Your personal experience*. These resources were mainly focused on the Mediterranean diet, based in most cases on locally and seasonally available products, or looking for a new combination of healthy foods to increase the interest of participants in healthy food and cooking. This information was based on 2 of the 3 essential phases of nutrition education: the motivational phase, which aimed to increase awareness and enhance motivation (*why* to make changes), and the action phase, which aimed to facilitate the ability to take action (*how* to make changes) [24]. Moreover, participants could self-assess their level of adherence to the Mediterranean diet through the 14-item Mediterranean Diet Adherence Screener (MEDAS) questionnaire (Multimedia Appendix 2) [25]. Multimedia Appendix 3 provided screenshots of the web page.

The information of the different areas of the web page was updated every week and was sequentially and automatically activated according to the number of weeks that each participant had been followed up in the study.

Recommended foods encompassed a total of 53 typical foods from the Mediterranean diet. Each factsheet provided an overview of the food, including the definition, the portion size, the recommended frequency of consumption, the nutritional value, health benefits, examples of how to include it in the diet, and 2 simple and delicious recipes.

The first entry of the *Menus* area was a week eating plan adapted to achieve the Mediterranean dietary pattern and a table that specified the recommended frequency of consumption for each group of food, which foods were included in the group, and the portion size. The following entries, 26 menus, provided examples to follow the eating guidelines outlined in the week eating plan. In general, each menu comprised 8 recipes based on crude and cooked vegetables, 3 recipes based on legumes, 3 recipes based on whole cereals, 3 recipes based on lean meats, 7 recipes based on fish, and 4 recipes based on eggs. Some menus specifically focused on a Mediterranean breakfast, healthy snacks, desserts based on fruit, and homemade healthy takeaway, among others. Each menu focused on one objective of the Mediterranean diet pattern and included a shopping list for one person for a week.

The area *News and Online resources* included a total of 71 news items and 7 web pages for the general public. In each news item, we provided our opinion, based on scientific evidence, about news regarding health and diet, taking into account that sometimes the information is confusing for the general population. The web pages were also related to diet and health and included blogs and web-based tools to increase adherence to the Mediterranean diet.

Practical tips were developed to facilitate the ability to dietary change. The tips used graphic images to calculate the hand-based portion size of different food groups, identify the seasonality of fruits and vegetables, guide healthy food shopping, guide how to eat healthy food in a restaurant, and identify the benefits of eating in family, among others. Recommendations using hand-based portion sizes were followed because they are considered to be an easy and acceptably accurate method for estimating portion sizes [26,27].

The *Mediterranean diet classroom* area consisted of videos about the theoretical and practical aspects of healthy nutrients, foods, and cooking. There were 24 videos related to theoretical aspects of nutrition, including information about nutrients (definition, classification, function, food source of the nutrient, and health effect) and food groups (definition, classification, health properties, portion size, and recommended frequency consumption). In addition, there were 12 videos with practical tips aimed to increase the adherence to the Mediterranean diet and 16 videos with novel but easily made recipes developed in collaboration with chefs of the Basque Culinary Center.

Finally, testimonials from volunteers of the PREDIMAR study were included in the *Your personal experience* area to motivate other volunteers to follow the Mediterranean diet.

Mobile App

The intervention program also included an Android and iPhone app that allowed participants to access the web page content in an easier manner (Multimedia Appendix 4). The mobile and tablet app was activated when 446 subjects (226 subjects in the intervention group) had begun the trial. Its availability was announced to all volunteers by email, and it included a pop-up on the web page and a paper sheet with the announcement in the print material modules.

Website and Mobile App Usage

Data on the frequency of access to the website and the mobile app were provided by the same web page to the dietitians, and this information was used to determine the level of engagement of each participant and to orient the intervention of the dietitian during the phone call with each participant during the follow-up. In these phone calls, participants were informed about the tools available on the web page and encouraged to use it every week.

Biweekly automated email notification was sent, announcing the updated information of each area according to the number of weeks that each participant had been followed up in the study.

Printed Material

Participants in the intervention group also received printed material across the time of the study. Once participants were assigned to the intervention group, they received a book about the traditional Mediterranean diet [28]. This book explains the main food habit objectives of the Mediterranean diet and provides more than 50 recipes.

Later in the first follow-up clinical visit (third month), participants in the intervention group received a binder with the first print module, which corresponds to the information of the first 8 weeks of the website. Subsequently, every 8 weeks, participants were sent 10 print modules with the information of the website.

A second book was provided to participants at the sixth month of the intervention [29]. This book shows the beneficial effects of healthy lifestyles, mainly focusing on the Mediterranean diet and its components.

Finally, in the 12th month clinical visit, a magnetic board was given to each volunteer of the intervention group. This board was a helpful tool to improve weekly eating plans.

Human Support (Phone Calls and Email)

The intervention began with a phone call from the dietitian once the patient had agreed to participate in the study and was randomly assigned to the intervention or control group. During this first phone call, the dietitian collected information about lifestyle, nutrition, and quality of life from all participants [25,30,31]. Those participants in the control group received only general information about the study, and no nutritional information was provided to them. Participants in the control group were informed that they would receive a second phone call after 1 year of follow-up to collect further information and a third phone call at the end of the study after 2 years of follow-up. They were informed that they would receive a gift when they visited their cardiologist at the end of the study. This gift was a book about the Mediterranean diet [28] and 3 L of EVOO, but they were not told this to minimize any nutritional intervention.

Every 3 months during the 2-year follow-up, participants in the intervention group were contacted by the dietitian by phone to complete the MEDAS questionnaire (Multimedia Appendix 2) and to conduct the nutritional education session [25]. The nutritional information collected in the food frequency questionnaire (FFQ) (Multimedia Appendix 5) and the updated information obtained with the MEDAS questionnaire were used by the dietitian to provide personalized nutritional information. The MEDAS questionnaire was designed to assess the level of compliance with the Mediterranean diet as well as to improve adherence to the Mediterranean diet [25]. Moreover, in each phone call, participants in the intervention group were asked about their use of the web page or mobile app and the printed material.

The aim of this tailored nutritional education was to increase the quality of the participants' diet according to the traditional Mediterranean diet and to adapt these general recommendations to the personal food preferences and lifestyles of each participant. Thus, volunteers with excessive body weight, with diabetes, or with any other disease related to nutrition, received specific recommendations by the dietitian to avoid contradictory information from other care professionals because in some cases, the intake of some foods is limited.

On the basis of behavioral literature showing the importance of continued contact during intervention [32], a specific section was designed on the website in which participants could ask during the study any questions regarding diet and health topics. Later, participants' questions were included anonymously in a *frequently asked questions* area.

EVOO

Each participant in the intervention group received 0.5 L of EVOO per week for free. The EVOO was provided at each clinical visit, and the aim was to encourage participants to consume at least four spoons of EVOO per day. As part of the Mediterranean diet intervention, participants were encouraged to use EVOO as the culinary fat in their homes.

Dietary Assessment

Trained registered dietitians collected data on food habits during the phone calls at baseline and at years 1 and 2 of follow-up. Adherence to the traditional Mediterranean diet was appraised by a validated 14-item MEDAS questionnaire (Multimedia Appendix 2) [25]. In the intervention group, this tool was used to assess the level of compliance with the intervention and as a key element to guide the personalized motivational interviews during the follow-up study every 3 months. Dietary intake was analyzed by a 147-item semiquantitative FFQ validated for the Spanish population (Multimedia Appendix 5) [30]. The FFQ included 9 frequency options for a specified usual portion size (never or less than once a month, 1-3 times a month, once a week, 2-4 times a week, 5-6 times a week, once a day, 2-3 times a day, 4-6 times a day, and more than 6 times a day). Energy and nutrient intake were calculated from Spanish food composition tables [33,34]. For the present analysis, changes in food consumption were assessed for 20 food groups: fruits, vegetables, refined olive oil, virgin olive oil (VOO), other fats than olive oil, whole grain cereals, refined cereals, dairy products, pulses, nuts, white fish, fatty fish, white meat, red meat, eggs, sweets, red wine, other wines than red wine, beer, and other alcoholic drinks (liquors, distilled beverages); and 9 nutrients, carbohydrates, protein, total fat, monounsaturated fatty acids (MUFAs), polyunsaturated fatty acids (PUFAs), saturated fatty acids (SFAs), omega-3, dietary fiber, and sodium.

Information on the hydroxytyrosol content in foods was obtained from the Phenol-Explorer database. When an item of the FFQ included more than one food, we used a weighted average according to the typical relative frequency of consumption in the Spanish population [35]. If this information was not available, the consumption of the foods included in the same item was equally divided. For recipes, polyphenol content was calculated according to the ingredients. The retention factors from the Phenol-Explorer database were applied to consider food cooking and processing to calculate the hydroxytyrosol content.

Other Measurements

At baseline, registered dietitians also collected information about sociodemographic characteristics (education level, civil status, and working status), smoking habit, and anthropometric measurements (self-reported weight and height). During this phone call, information about the physical activity was also collected with a physical activity questionnaire validated for the Spanish population [31].

In clinical visits, the cardiologists collected AF-related variables, complications related to catheter ablation, presence of concomitant chronic diseases, and changes in medication related to arrhythmia, among others. Electrocardiographic monitoring was performed at each visit [20].

Statistical Analyses

For this study, we used the PREDIMAR database generated until September 2020, including 1- and 2-year follow-up data. Quantitative variables were expressed as means and SDs, whereas categorical variables were described as number and percentages (n [%]). The Student *t* test for continuous variables and the chi-square test for categorical variables were applied to test differences in baseline characteristics between the intervention groups. Mixed effects linear models were used to assess changes in nutritional variables from baseline to 12- and 24-month follow-up visits. A 2-level mixed linear model with random intercepts at the recruitment center and participant was fitted. The analyses were performed using STATA (v 12.0, StataCorp LP). The significance level (2-tailed) was set at *P* values lower than .05.

Results

Baseline Characteristics

Between March 2017 and January 2020, 720 patients started the intervention (Figure 1). Among them, 549 were men (76%) and 171 were women (24%), and the mean age was 59.7 years (SD 10.7). Table 1 shows the demographic, anthropometric, and lifestyle baseline characteristics of participants according to the randomized groups (intervention or control). No significant differences were found for age, sex, type of AF, BMI, educational level, civil and working status, smoking habit, physical activity, and protein intake between intervention groups. Meanwhile, significant differences were observed in dietary habits. Participants in the intervention group had a lower intake of energy and carbohydrates and a higher intake of fat than participants in the control group. Moreover, participants in the intervention group had a higher adherence to the Mediterranean diet. These differences could be because of social desirability bias, although the magnitude of the differences between groups was small, and they could be interpreted in the light of the large power and sample size of the study.

Table 1. Baseline characteristics of the participants in the prevention of recurrent arrythmias with Mediterranean diet trial.

Characteristics	Intervention (n=365)	Control (n=355)	*P* value
Age, years, mean (SD)	59.9 (10.5)	59.6 (10.9)	.77
Sex, n (%)			**.41**
Men	283 (77.5)	266 (74.9)	
Women	82 (22.5)	89 (25.1)	
Type of AF[a], n (%)			**.94**
Persistent	147 (40.3)	142 (40.0)	
Paroxysmal	218 (59.7)	213 (60.0)	
BMI (kg/m^2)	27.8 (4.0)	27.8 (4.3)	.92
Education, n (%)			**.47**
Secondary or less	198 (54.2)	202 (56.9)	
University	167 (45.7)	153 (43.1)	
Civil status, n (%)			**.63**
Single	24 (6.6)	27 (7.6)	
Married	278 (76.2)	275 (77.5)	
Others	63 (17.3)	53 (14.9)	
Working status, n (%)			**.80**
Working	199 (54.5)	191 (53.8)	
Retired	139 (38.1)	133 (37.5)	
Others	27 (7.4)	31 (8.7)	
Smoking status, n (%)			**.39**
Never	139 (38.1)	121 (34.1)	
Former	206 (56.4)	208 (59.6)	
Current	20 (5.5)	26 (7.3)	
Physical activity (MET[b]-hours/week)	33.4 (20.7)	34.1 (22.1)	.67
MEDAS[c] score (14 items)	7.8 (2.0)	7.2 (2.0)	<.001
Energy intake (kcal/day)	2396 (670)	2527 (813)	.01
Carbohydrate intake g/day)	251.5 (5.2)	275.7 (5.6)	.002
Protein intake (g/day)	94.3 (1.3)	99.9 (1.4)	.002
Fat intake (g/day)	104.8 (1.5)	107.6 (2.0)	.21

[a] AF: atrial fibrillation.

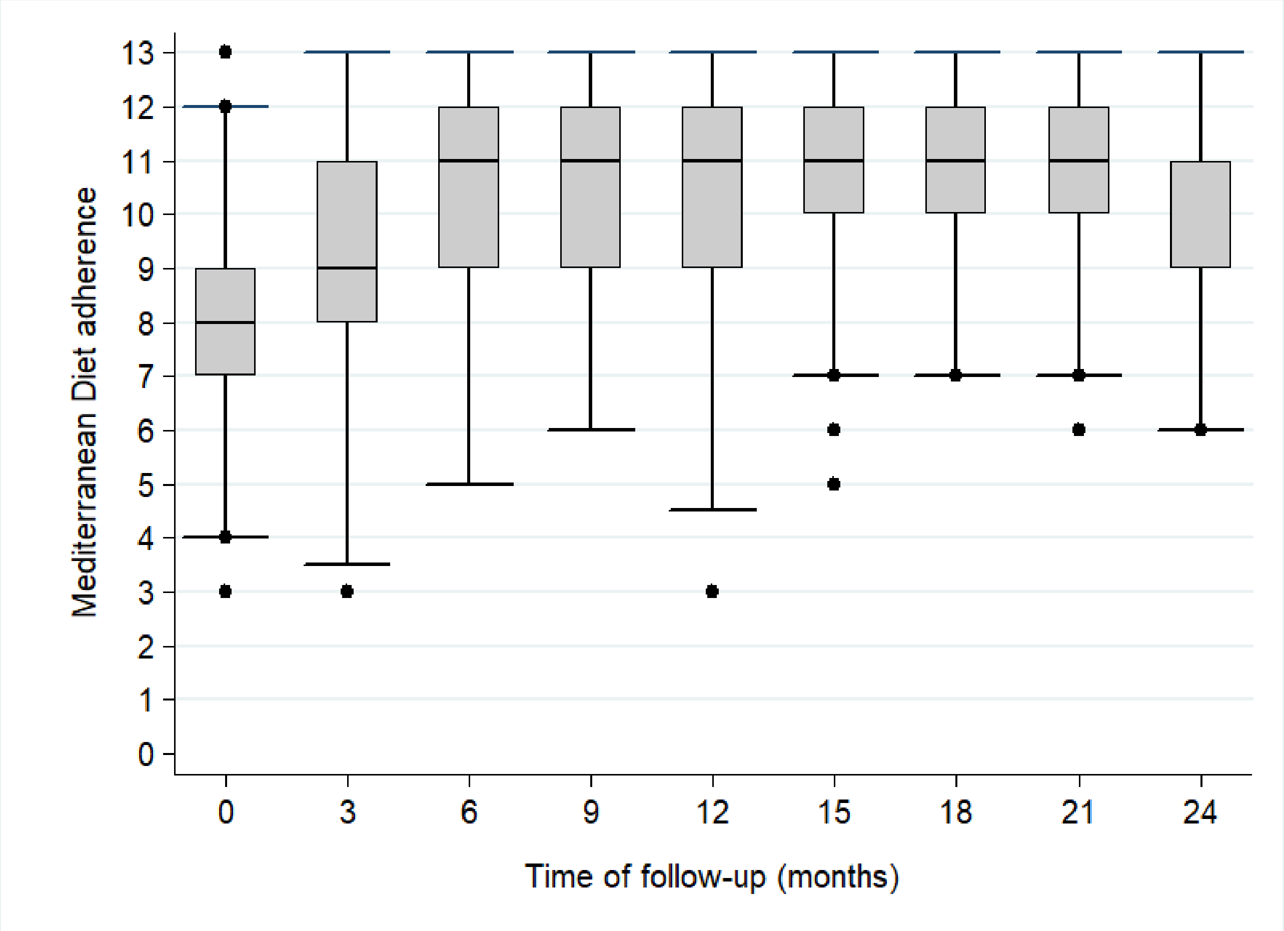
[b] MET: metabolic equivalent.

[c] MEDAS: Mediterranean Diet Adherence Screener.

Mediterranean Diet Adherence

After 12 and 24 months of follow-up, an increase in the adherence to the Mediterranean diet was observed in both groups (Multimedia Appendix 6). The mean (95% CI) MEDAS score was 7.8 (7.6-8.0) at baseline, 10.2 (10.0-10.4) at 12 months (increase, 2.4 [2.2-2.6]) and 10.2 (10.0-10.4) at 24 months (increase, 2.4 [2.1-2.6]) in the intervention group; and 7.2 (7.0-7.4) at baseline, 7.5 (7.2-7.7) at 12 months (increase, 0.3 [0.03-0.5]), and 7.9 (7.6-8.2) at 24 months (increase, 0.6 [0.4-0.9]) in the control group. Thus, the increase in Mediterranean diet adherence was higher in the intervention than in the control group at 12 months (between-group difference 2.1, 95% CI 1.8-2.4, *P*<.001) and 24 months (between-group difference 1.8, 95% CI 1.4-2.1; *P*<.001) of follow-up. Figure 2 shows the adherence to the Mediterranean diet for each 3-month follow-up visit among participants of the intervention group. The median score of the adherence to the Mediterranean diet increased gradually until the 6-month follow-up visit. After that, the median adherence was maintained until the 21-month follow-up phone call.

Figure 2. Adherence to the Mediterranean diet among participants in the intervention group in each follow-up phone call.



Food Group Consumption

As intended, the Mediterranean diet intervention group showed a significant improvement in the consumption of vegetables, fruits, whole grain cereals, pulses, nuts, white fish, fatty fish, white meat, and VOO compared with the control group at 1-year follow-up (Table 2). Moreover, the intervention group reduced the consumption of refined olive oil and unhealthy foods, refined cereals, red and processed meat, and sweets compared with the control group. No significant differences between groups were observed for the change in the consumption of dairy products, eggs, other fats than olive oil, wine, beer, and other alcoholic drinks.

Table 2. Baseline food groups consumption and changes by randomized treatment group at 12- and 24-month follow-up visits of participants in the prevention of recurrent arrythmias with Mediterranean diet trial.

Food groups	Group intervention		Between group difference[a], mean (95% CI)	*P* value[b]
	Intervention, mean (95% CI)	Control, mean (95% CI)		
Vegetables (g/day)				
Baseline	217.3 (208.1 to 226.5)	228.2 (217.3 to 239.1)	N/A[c]	N/A
1 year	**282.6 (269.2 to 295.9)**	**251.6 (237.7 to 265.5)**	**N/A**	**N/A**
1-year change	65.2 (52.1 to 78.4)	23.4 (9.8 to 37.0)	41.9 (23.0 to 60.8)	<.001
2 years	**250.4 (238.3 to 262.5)**	**280.8 (265.0 to 296.7)**	**N/A**	**N/A**
2 years change	33.1 (19.8 to 46.3)	52.6 (36.7 to 68.6)	−19.5 (−40.3 to 1.2)	.06
Fruits (g/day)				
Baseline	313.7 (293.2 to 334.2)	313.7 (291.6 to 335.9)	N/A	N/A
1 year	**464.9 (441.4 to 488.4)**	**327.6 (304.9 to 350.3)**	**N/A**	**N/A**
1-year change	151.2 (127.6 to 174.7)	13.9 (−6.8 to 34.6)	137.3 (105.9 to 168.7)	<.001
2 years	**439.5 (414.6 to 464.3)**	**304.3 (274.8 to 333.7)**	**N/A**	**N/A**
2 years change	125.7 (99.8 to 151.7)	−9.5 (−40.3 to 21.3)	135.2 (94.9 to 175.5)	<.001
Refined cereals (g/day)				
Baseline	122.8 (110.4 to 135.2)	131.8 (119.8 to 143.9)	N/A	N/A
1 year	**66.3 (57.8 to 74.8)**	**111.4 (99.9 to 122.9)**	**N/A**	**N/A**
1-year change	−56.5 (−67.8 to −45.2)	−20.4 (−32.0 to −8.8)	−36.1 (−52.3 to −19.9)	<.001
2 years	**56.2 (46.8 to 65.6)**	**108.3 (94.8 to 121.8)**	**N/A**	**N/A**
2 years change	−66.6 (−80.6 to −52.6)	−23.5 (−38.1 to −9.0)	−43.0 (−63.2 to −22.9)	<.001
Whole cereals (g/day)				
Baseline	40.3 (33.5 to 47.2)	36.2 (29.3 to 43.1)	N/A	N/A
1 year	**50.5 (43.7 to 57.3)**	**28.6 (22.7 to 34.4)**	**N/A**	**N/A**
1-year change	10.1 (1.9 to 18.4)	−7.7 (−15.3 to 0.00)	17.8 (6.5 to 29.1)	.002
2 years	**47.5 (40.1 to 54.8)**	**23.0 (15.0 to 31.0)**	**N/A**	**N/A**
2 years change	7.1 (−1.1 to 15.3)	−13.2 (−22.8 to −3.7)	20.3 (7.8 to 32.9)	.002
Pulses (g/week)				
Baseline	120.7 (112.9 to 128.4)	131.5 (122.5 to 140.4)	N/A	N/A
1 year	**186.3 (175.6 to 197.1)**	**145.5 (133.8 to 157.2)**	**N/A**	**N/A**
1-year change	65.6 (53.9 to 77.4)	14.0 (0.4 to 27.7)	51.6 (33.6 to 69.7)	<.001
2 years	**175.5 (168.4 to 182.6)**	**133.5 (121.5 to 145.4)**	**N/A**	**N/A**
2 years change	54.9 (44.6 to 65.1)	2.0 (−12.3 to 16.2)	52.9 (35.3 to 70.4)	<.001
Nuts (g/week)				
Baseline	133.9 (119.7 to 148.0)	88.9 (78.7 to 99.1)	N/A	N/A
1 year	**200.4 (186.3 to 214.5)**	**103.0 (87.0 to 119.1)**	**N/A**	**N/A**
1-year change	66.5 (48.5 to 84.5)	14.2 (−0.23 to 28.6)	52.4 (29.3 to 75.4)	<.001
2 years	**193.8 (177.0 to 210.5)**	**101.8 (87.1 to 116.5)**	**N/A**	**N/A**
2 years change	59.9 (40.1 to 79.7)	12.9 (−2.7 to 28.6)	47.0 (21.7 to 72.2)	<.001
Dairy products (g/week)				
Baseline	292.4 (272.5 to 312.3)	333.1 (309.4 to 356.9)	N/A	N/A
1 year	**301.5 (279.2 to 323.8)**	**351.3 (324.5 to 378.1)**	**N/A**	**N/A**
1-year change	9.1 (−11.3 to 29.5)	18.2 (−7.7 to 44.0)	−9.1 (−42.0 to 23.8)	.59

Food groups	Group intervention		Between group difference[a], mean (95% CI)	*P* value[b]
	Intervention, mean (95% CI)	Control, mean (95% CI)		
2 years	**308.1 (283.0 to 333.3)**	**379.4 (349.7 to 409.1)**	**N/A**	**N/A**
2 years change	15.7 (−7.6 to 39.1)	46.3 (18.8 to 73.8)	−30.5 (−66.6 to 5.5)	.10
White fish (g/week)				
Baseline	443.1 (415.5 to 470.7)	483.4 (453.9 to 513.0)	N/A	N/A
1 year	**548.7 (521.6 to 575.8)**	**482.1 (448.0 to 516.2)**	**N/A**	**N/A**
1-year change	105.6 (74.1 to 137.1)	−1.3 (−32.9 to 30.3)	106.9 (62.3 to 151.6)	<.001
2 years	**528.0 (499.8 to 556.2)**	**450.4 (414.3 to 486.6)**	**N/A**	**N/A**
2 years change	84.9 (49.7 to 120.2)	−33.0 (−70.4 to 4.4)	117.9 (66.5 to 169.3)	<.001
Fatty fish (g/week)				
Baseline	210.6 (195.4 to 225.9)	222.0 (204.9 to 239.1)	N/A	N/A
1 year	**315.3 (290.8 to 339.7)**	**241.1 (222.2 to 260.1)**	**N/A**	**N/A**
1-year change	104.6 (79.6 to 129.6)	19.1 (−1.1 to 39.3)	85.5 (53.3 to 117.6)	<.001
2 years	**280.6 (259.0 to 302.2)**	**231.7 (206.4 to 257.0)**	**N/A**	**N/A**
2 years change	70.0 (46.1 to 93.8)	9.7 (−13.8 to 33.2)	60.3 (26.8 to 93.7)	<.001
White meat (g/week)				
Baseline	436.6 (414.5 to 458.7)	392.3 (369.7 to 414.8)	N/A	N/A
1 year	**458.8 (435.3 to 482.3)**	**368.7 (345.6 to 391.7)**	**N/A**	**N/A**
1-year change	22.2 (−0.7 to 45.1)	−23.6 (−46.6 to −0.6)	45.8 (13.3 to 78.3)	.006
2 years	**466.0 (440.1 to 491.9)**	**363.0 (331.0 to 395.1)**	**N/A**	**N/A**
2 years change	29.4 (2.4 to 56.3)	−29.2 (−62.3 to 3.9)	58.6 (15.9 to 101.2)	.007
Red and processed meat (g/week)				
Baseline	580.6 (542.3 to 619.0)	655.8 (618.1 to 693.5)	N/A	N/A
1 year	**444.2 (412.0 to 476.4)**	**634.4 (587.9 to 681.0)**	**N/A**	**N/A**
1-year change	−136.4 (−174.3 to −98.6)	−21.4 (−63.5 to 20.8)	−115.1 (−171.8 to −58.4)	<.001
2 years	**408.4 (370.9 to 445.9)**	**603.5 (559.8 to 647.2)**	**N/A**	**N/A**
2 years change	−172.2 (−213.6 to −130.9)	−52.3 (−90.9 to −13.7)	−119.9 (−176.5 to −63.3)	<.001
Eggs (g/week)				
Baseline	29.3 (27.1 to 31.5)	28.3 (26.3 to 30.3)	N/A	N/A
1 year	**29.8 (27.6 to 31.9)**	**30.6 (27.8 to 33.3)**	**N/A**	**N/A**
1-year change	0.4 (−1.5 to 2.4)	2.2 (−0.3 to 4.8)	−1.8 (−5.0 to 1.4)	.27
2 years	**30.0 (27.7 to 32.3)**	**32.2 (29.0 to 35.3)**	**N/A**	**N/A**
2 years change	0.7 (−1.5 to 2.8)	3.9 (1.2 to 6.6)	−3.2 (−6.7 to 0.2)	.07
Refined olive oil (g/day)				
Baseline	2.9 (2.0 to 3.8)	4.0 (2.9 to 5.2)	N/A	N/A
1 year	**0.3 (0.1 to 0.6)**	**3.9 (2.6 to 5.1)**	**N/A**	**N/A**
1-year change	−2.7 (−3.6 to −1.7)	−0.2 (−1.5 to 1.1)	−2.5 (−4.1 to −0.9)	.003
2 years	**0.5 (0.07 to 1.0)**	**5.2 (3.4 to 7.0)**	**N/A**	**N/A**
2 years change	−2.4 (−3.3 to −1.5)	1.2 (−0.7 to 3.0)	−3.6 (−5.6 to −1.5)	.001
Virgin olive oil (g/day)				
Baseline	34.5 (32.8 to 36.2)	30.5 (28.7 to 32.3)	N/A	N/A
1 year	**45.0 (43.5 to 46.5)**	**28.3 (26.3 to 30.3)**	**N/A**	**N/A**

Food groups	Group intervention		Between group difference[a], mean (95% CI)	*P* value[b]
	Intervention, mean (95% CI)	Control, mean (95% CI)		
1-year change	10.5 (8.5 to 12.4)	−2.2 (−4.4 to −0.0)	12.7 (9.7 to 15.7)	<.001
2 years	**43.3 (41.3 to 45.4)**	**27.9 (25.3 to 30.5)**	**N/A**	**N/A**
2 years change	8.8 (6.3 to 11.3)	−2.6 (−5.5 to 0.3)	11.4 (7.6 to 15.3)	<.001
Other fats than olive oil (g/day)				
Baseline	2.0 (1.5 to 2.4)	3.0 (2.2 to 3.7)	N/A	N/A
1 year	**0.7 (0.4 to 1.0)**	**2.2 (1.6 to 2.7)**	**N/A**	**N/A**
1-year change	−1.3 (−1.7 to −0.9)	−0.8 (−1.5 to −0.04)	−0.5 (−1.3 to 0.3)	.25
2 years	**0.7 (0.4 to 1.1)**	**2.9 (1.8 to 4.0)**	**N/A**	**N/A**
2 years change	−1.2 (−1.7 to −0.7)	−0.02 (−1.1 to 1.1)	−1.2 (−2.4 to 0.02)	.05
Sweets (g/week)				
Baseline	217.7 (188.7 to 246.6)	392.9 (335.2 to 450.7)	N/A	N/A
1 year	**174.6 (146.3 to 203.0)**	**443.5 (384.7 to 502.3)**	**N/A**	**N/A**
1-year change	−43.0 (−79.3 to −6.7)	50.5 (−25.3 to 126.4)	−93.5 (−177.6 to −9.5)	.03
2 years	**135.8 (115.7 to 155.9)**	**487.8 (419.3 to 556.3)**	**N/A**	**N/A**
2 years change	−81.9 (−113.4 to −50.4)	94.9 (17.9 to 171.9)	−176.8 (−260.0 to −93.6)	<.001
Red wine (g of alcohol/day)				
Baseline	5.3 (4.3 to 6.2)	4.4 (3.5 to 5.3)	N/A	N/A
1 year	**6.3 (5.2 to 7.4)**	**4.9 (3.9 to 6.0)**	**N/A**	**N/A**
1-year change	1.0 (0.1 to 1.9)	0.5 (−0.2 to 1.3)	0.5 (−0.7 to 1.7)	.42
2 years	**6.4 (5.1 to 7.6)**	**4.7 (3.7 to 5.8)**	**N/A**	**N/A**
2 years change	1.1 (0.04 to 2.2)	0.3 (−0.5 to 1.2)	0.8 (−0.6 to 2.1)	.27
Other wines than red wine (g of alcohol/day)				
Baseline	0.7 (0.2 to 1.1)	0.4 (0.2 to 0.5)	N/A	N/A
1 year	**0.9 (0.5 to 1.4)**	**0.8 (0.4 to 1.2)**	**N/A**	**N/A**
1-year change	0.2 (−0.1 to 0.6)	0.5 (0.1 to 0.9)	−0.2 (−0.7 to 0.2)	.33
2 years	**0.7 (0.1 to 1.2)**	**0.8 (0.4 to 1.2)**	**N/A**	**N/A**
2 years change	−0.01 (−0.6 to 0.5)	0.4 (0.07 to 0.8)	−0.4 (−1.1 to 0.2)	.18
Beer (g of alcohol/day)				
Baseline	3.1 (2.5 to 3.8)	2.7 (2.1 to 3.4)	N/A	N/A
1 year	**2.4 (1.8 to 3.0)**	**3.0 (2.3 to 3.6)**	**N/A**	**N/A**
1-year change	−0.7 (−1.3 to −0.07)	0.2 (−0.5 to 1.0)	−0.9 (−1.9 to 0.04)	.06
2 years	**2.5 (1.9 to 3.1)**	**3.3 (2.5 to 4.1)**	**N/A**	**N/A**
2 years change	−0.6 (−1.2 to −0.03)	0.5 (−0.4 to 1.5)	−1.2 (−2.3 to −0.06)	.04
Other alcoholic drinks (g of alcohol/day)				
Baseline	0.7 (0.4 to 1.0)	0.5 (0.4 to 0.7)	N/A	N/A
1 year	**0.6 (0.3 to 0.9)**	**0.8 (0.6 to 1.1)**	**N/A**	**N/A**
1-year change	−0.09 (−0.5 to 0.3)	0.3 (0.02 to 0.6)	−0.4 (−0.9 to 0.09)	.11
2 years	**0.4 (0.3 to 0.6)**	**0.9 (0.4 to 1.4)**	**N/A**	**N/A**
2 years change	−0.2 (−0.6 to 0.05)	0.4 (−0.09 to 0.8)	−0.6 (−1.2 to −0.07)	.03

[a] Calculated using mixed-effect models with center as random factor.

[b] *P* value between group intervention difference.

[c] N/A: not applicable.

At 2 years of intervention, between-group differences were sustained except for the consumption of vegetables, beer, and other alcoholic drinks (liquors and distilled beverages). There was a significant increase in the consumption of vegetables within the intervention group, although the difference in changes between the intervention groups was not statistically significant. The intervention group reduced the consumption of beer and other alcoholic drinks (liquors and distilled beverages) compared with the control group after 2 years of follow-up, but not during the first year of follow-up.

Energy and Nutrient Intake

Consistent with changes in consumption of food groups associated with the Mediterranean diet, significant between-group differences were observed for increased intake of fat, MUFA, PUFA, omega-3, and fiber, for the intervention group versus the control group at 1 year of the intervention (Table 3). After a 2-year follow-up, there were no significant differences between groups in the intake of fat, MUFA, and PUFA. The intervention group showed a decrease in the intake of sodium at 1 and 2 years of follow-up and SFA and carbohydrates after 2 years when compared with the control group.

Table 3. Baseline nutrient intake and changes by randomized treatment group at 12- and 24-month follow-up visits of participants in the prevention of recurrent arrythmias with Mediterranean diet trial.

Energy or nutrient	Group intervention		Between group difference[a], mean (95% CI)	*P* value[b]
	Intervention, mean (95% CI)	Control, mean (95% CI)		
Energy (kcal/day)				
Baseline	2396 (2328 to 2465)	2527 (2443 to 2612)	N/A[c]	N/A
1 year	**2377 (2320 to 2435)**	**2466 (2379 to 2552)**	**N/A**	**N/A**
1-year change	−9.1 (−85.9 to 47.7)	−61.6 (−154.8 to 31.7)	42.5 (−71.2 to 157.2)	.47
2 years	**2250 (2191 to 2309)**	**2496 (2388 to 2604)**	**N/A**	**N/A**
2 years change	−1467 (−224 to -68.1)	−31.5 (−133.5 to 70.4)	−114.6 (−243.0 to 13.7)	.08
Fat (g/day)				
Baseline	104.8 (101.8 to 107.8)	107.6 (103.7 to 111.5)	N/A	N/A
1 year	**110.4 (107.6 to 113.3)**	**106.0 (102.0 to 109.9)**	**N/A**	**N/A**
1-year change	5.6 (2.2 to 9.1)	−1.6 (−6.0 to 2.7)	7.3 (1.7 to 12.8)	.01
2 years	**106.0 (102.9 to 109.0)**	**110.0 (105.0 to 115.0)**	**N/A**	**N/A**
2 years change	1.2 (−2.6 to 4.9)	2.4 (−2.6 to 7.5)	−1.3 (−7.6 to 5.1)	.70
MUFA[d] (g/day)				
Baseline	54.2 (52.6 to 55.9)	54.6 (52.5 to 56.8)	N/A	N/A
1 year	**59.9 (58.2 to 61.7)**	**54.5 (52.2 to 56.9)**	**N/A**	**N/A**
1-year change	5.7 (3.5 to 7.8)	−0.1 (−2.7 to 2.5)	5.8 (2.5 to 9.1)	.001
2 years	**57.3 (55.3 to 59.3)**	**57.0 (54.1 to 59.9)**	**N/A**	**N/A**
2 years change	3.0 (0.7 to 5.4)	2.3 (−0.8 to 5.4)	0.7 (−3.2 to 4.6)	.72
PUFA[e] (g/day)				
Baseline	18.7 (18.0 to 19.3)	18.7 (17.9 to 19.5)	N/A	N/A
1 year	**20.5 (19.7 to 21.2)**	**18.3 (17.5 to 19.2)**	**N/A**	**N/A**
1-year change	1.8 (0.9 to 2.7)	−0.4 (−1.3 to 0.6)	2.2 (0.9 to 3.5)	.001
2 years	**19.7 (18.8 to 20.5)**	**18.6 (17.7 to 19.6)**	**N/A**	**N/A**
2 years change	1.0 (0.02 to 2.0)	−0.06 (−1.1 to 1.0)	1.1 (−0.4 to 2.5)	.15
SFA[f] (g/day)				
Baseline	26.2 (25.3 to 27.2)	27.8 (26.6 to 28.9)	N/A	N/A
1 year	**24.6 (23.8 to 25.3)**	**27.2 (26.0 to 28.3)**	**N/A**	**N/A**
1-year change	−1.7 (−2.5 to −0.8)	−0.6 (−1.7 to 0.5)	−1.1 (−2.5 to 0.3)	.14
2 years	**23.7 (22.9 to 24.6)**	**28.2 (26.8 to 29.7)**	**N/A**	**N/A**
2 years change	−2.5 (−3.5 to −1.5)	0.5 (−0.8 to 1.7)	−3.0 (−4.6 to −1.3)	<.001
Omega 3 (mg/day)				
Baseline	0.8 (0.8 to 0.8)	0.9 (0.8 to 0.9)	N/A	N/A
1 year	**1.1 (1.0 to 1.2)**	**0.9 (0.9 to 1.0)**	**N/A**	**N/A**
1-year change	0.3 (0.2 to 0.4)	0.05 (−0.01 to 0.1)	0.3 (0.2 to 0.4)	<.001
2 years	**1.0 (0.9 to 1.1)**	**0.9 (0.8 to 1.0)**	**N/A**	**N/A**
2 years change	0.2 (0.1 to 0.3)	0.01 (−0.05 to 0.08)	0.2 (0.09 to 0.3)	<.001
Protein (g/day)				
Baseline	94.3 (91.8 to 96.8)	99.9 (97.2 to 102.5)	N/A	N/A
1 year	**96.3 (93.9 to 98.7)**	**98.8 (95.8 to 101.8)**	**N/A**	**N/A**

Energy or nutrient	Group intervention		Between group difference[a], mean (95% CI)	*P* value[b]
	Intervention, mean (95% CI)	Control, mean (95% CI)		
1-year change	2.0 (−0.4 to 4.4)	−1.1 (−3.9 to 1.8)	3.1 (−0.6 to 6.8)	.10
2 years	**91.3 (88.8 to 93.8)**	**97.7 (94.0 to 101.4)**	**N/A**	**N/A**
2 years change	−3.0 (−5.6 to −0.3)	-2.2 (−5.4 to 1.1)	−0.8 (−5.0 to 3.4)	.70
Carbohydrates (g/day)				
Baseline	251.5 (241.4 to 261.6)	275.7 (264.1 to 287.2)	N/A	N/A
1 year	**231.3 (223.5 to 239.1)**	**262.0 (250.5 to 273.5)**	**N/A**	**N/A**
1-year change	−20.2 (−29.9 to −10.5)	−13.7 (−26.6 to −0.7)	−6.6 (−22.7 to 9.6)	.43
2 years	**215.1 (206.8 to 223.5)**	**260.3 (246.1 to 274.5)**	**N/A**	**N/A**
2 years change	−36.4 (−47.7 to −25.1)	−15.3 (−29.4 to −1.3)	−21.0 (−39.1 to −3.0)	.02
Fiber (g/day)				
Baseline	24.1 (23.2 to 25.0)	25.1 (24.1 to 26.2)	N/A	N/A
1 year	**30.2 (29.2 to 31.2)**	**25.7 (24.5 to 26.9)**	**N/A**	**N/A**
1-year change	6.1 (5.0 to 7.2)	0.5 (−0.8 to 1.9)	5.6 (3.8 to 7.3)	<.001
2 years	**28.2 (27.1 to 29.3)**	**25.1 (23.8 to 26.5)**	**N/A**	**N/A**
2 years change	4.1 (2.9 to 5.2)	−0.02 (−1.5 to 1.4)	4.1 (2.2 to 6.0)	<.001
Sodium (mg/day)				
Baseline	3390 (3267 to 3512)	3523 (3398 to 3648)	N/A	N/A
1 year	**2831 (2719 to 2943)**	**3254 (3119 to 3390)**	**N/A**	**N/A**
1-year change	−558.4 (−684.3 to −432.6)	−268.9 (−403.9 to −133.9)	−289.5 (−474.1 to −105.0)	.002
2 years	**2586 (2464 to 2709)**	**3214 (3048 to 3380)**	**N/A**	**N/A**
2 years change	−803.6 (-943.1 to −664.1)	−308.9 (−472.0 to −145.7)	−494.7 (−709.4 to −280.1)	<.001

[a] Calculated using mixed-effect models with center as random factor.

[b] *P* value between group intervention difference.

[c] N/A: not applicable.

[d] MUFA: monounsaturated fatty acid.

[e] PUFA: polyunsaturated fatty acid.

[f] SFA: saturated fatty acid.

Hydroxytyrosol Intake

Finally, regarding total hydroxytyrosol intake, no significant differences between groups at the 1- and 2-year follow-up visits were found. However, when we studied the intake of hydroxytyrosol derived from olive oil (one of the two main sources of this polyphenol in the diet), we observed that the intervention group increased the intake of hydroxytyrosol from olive oil after 1 and 2 years of follow-up, compared with the control group. Meanwhile, the intake of hydroxytyrosol derived from wine (the other main source of hydroxytyrosol of the diet) did not change at 1- and 2-year follow-up visits between groups.

Discussion

Principal Findings

First, this work describes in detail the nutritional intervention of the PREDIMAR trial, which is, to our knowledge, the first remote dietary intervention based on the Mediterranean diet specifically designed for patients with AF treated with catheter ablation. Second, our results demonstrate that a remote nutritional intervention is a useful tool kit to improve the quality of the diet according to the goals of the Mediterranean diet.

Comparison With Prior Work

Although nutritional interventions are typically derived face to face, at present, remote nutritional interventions (web page, mobile phone app, email, text messaging, and phone calls) are becoming more frequent. This could be in part because of the increase in the use of the internet and other technological resources as well as the high economic costs of traditional interventions. As far as we know, our research is unique in that it is a remote nutritional intervention using different behavioral change strategies including a web page and mobile app, printed material, and personalized advice by phone call and email. Remotely, intervention permits overcoming the barriers of in-person interventions such as lack of staff and institutional resources to reach a large number of participants, and that the participants have to attend to the onsite meetings (group or individual meetings), which in turn require substantial organizational skills [36]. On the other hand, the provision of personalized advice is associated with a greater change in dietary habits [36]. However, Hutchesson et al [37] reported that a web page–based intervention combined with comprehensive personalized feedback reports conducted similar improvements in dietary intake compared with the web-page-based intervention group. Therefore, we hypothesized that combining different nutritional intervention remotely tools could improve in a better manner the impact of the intervention. In fact, a previous systematic review concluded that using multiple modes of communication increases the effectiveness of remote intervention [22]. Moreover, it is widely acknowledged that web-based interventions provide effective changes in physical activity, eating behaviors, and weight loss [21-23]. To maintain engagement with the nutritional intervention, biweekly automated email notifications were sent to each participant. Moreover, we also used printed material to overcome potential barriers to internet access, especially among older participants [38,39]. In this context, the ManUp study concluded that an information technology-based (web and mobile) intervention was as effective as a print-based intervention [38].

Participants' baseline scores showed that they had a reasonably good Mediterranean-style food pattern, according to previous studies developed in the Mediterranean area with patients at risk or with any cardiovascular disease [40-43]. The difference between groups in the MEDAS score at baseline is consistent with previous nutritional studies [40,41,43] and could be due in part to social desirability bias, which means that participants in the intervention group tended to report higher Mediterranean diet adherence. After 2 years of follow-up, not only the intervention group, but also the control group increased the adherence to the Mediterranean diet. In this sense, it is possible that participants in the control group were aware of the relationship between food habits and cardiovascular health and thus were motivated to change their dietary habits after catheter ablation. However, the improvement in the adherence to the Mediterranean diet was significantly higher in the intervention group, and the magnitude of the observed changes was similar to that in face-to-face intervention studies. As an example, in the PREDIMED trial, participants randomized to the Mediterranean diet group showed an increase of 2.3 points in the MEDAS score after the 12-month intervention period [40]. A more recent example is the PREDIMED-plus trial, in which the mean increase in adherence to the Mediterranean diet was 2.1 points in the Mediterranean diet group (control group) after 1 year of follow-up [42]. Moreover, our results are consistent with those of previous online intervention studies. Recio-Rodriguez et al [44] and Choi et al [45] demonstrated, using the MEDAS questionnaire, the efficacy of a mobile application intervention in promoting the Mediterranean diet. Neither study found a significant difference with the Mediterranean diet changes found in the face-to-face intervention group.

We observed a higher increase in the consumption of plant-based foods (fruits, vegetables, whole grain cereals, olive oil, pulses, and nuts) and fish in the intervention group than in the control group during the intervention period. Similar results were found in the PREDIMED study after 1 year of face-to-face intervention [40,41]. Moreover, previous online intervention studies found an increase in the intake of fruit and vegetables [22,46]. Other authors have reported a significant increase in the consumption of total fat and different fatty acids after different online interventions [47,48]. Finally, Hutchesson et al [37] demonstrated that participants of a web-based weight loss program significantly reduced energy-dense, nutrient-poor foods.

It is evident that the intervention group adopted healthier behaviors during the first year of follow-up. However, in general, dietary habits were sustained in the longer term, and even a high decrease in the intake of red and processed meat, refined cereals, and sweets was observed at 2 years of follow-up among participants of the intervention group. These findings are consistent with those reported in the literature [49,50].

In line with previous intervention studies, we found an increase of 1 serving per day of VOO (including extra virgin) and a reduction of refined olive oil in the intervention group compared with the control group [42]. The magnitude of the change in the consumption of VOO may reflect a small increase in the individual effect, but a mean change of this magnitude may have a great impact at the population level. In this sense, a recent study demonstrated that, compared with nonconsumers or low consumers (0-<4.5 g per day), those with a high consumption of olive oil (>7 g per day) had 14% and 18% lower risk of cardiovascular and coronary heart disease, respectively [51]. On the other hand, previous research suggests that free provision of key food items is an effective strategy to increase adherence to the prescribed interventions [40,42]. Therefore, we hypothesized that the remote intervention combining general nutritional information and personalized advice, together with the free provision of EVOO, may be very effective in increasing adherence to a Mediterranean dietary pattern.

Limitations

There are several strengths and limitations of this study that should be considered when interpreting the results. First, these findings are based on preliminary analyses within the context of an ongoing randomized controlled trial, and it is unknown whether and how these results may be related to beneficial health outcomes. Second, the results of the nutritional intervention may not be applicable to the general population for 2 main reasons. On the one hand, the population of this study was patients with AF, and therefore they could be worried about their health. On the other hand, the free provision of EVOO, which could be a strength of our study, can also represent a barrier because of the high cost of this product. Third, although the clinical providers were blinded to the allocation group, the dietitians were not blinded. Fourth, we used a self-reported FFQ instead of objective instruments, such as biomarkers. Recall bias, social desirability bias, and other potential reporting biases may have affected the results. However, the FFQ has been previously validated and is suitable for repeatedly ranking people according to their food and nutrient intake [30]. Fifth, the FFQ did not differ between the consumption of VOO and EVOO. However, in Spain, the consumption of EVOO outweighs the consumption of VOO (39% vs 9%) [52] and both have similar contents of MUFA and polyphenols [34,53,54]. A higher content of MUFA and polyphenols explains the attributed health effects of olive oil; therefore, both VOO and EVOO could have a similar beneficial effect on AF. Sixth, self-reported use of nutritional intervention tools (website, app, and printed materials) may not fully reflect the completion of health education. However, periodical phone calls from the dietitian were used as a monitor system to assess and meet the educational needs of each participant in the intervention group. Finally, we acknowledge that our results do not provide evidence to indicate that a remote intervention is more effective than an in-person intervention because this study did not use a control group with face-to-face intervention.

Conclusions

We found that a multifaceted remote nutritional intervention seems to be effective in increasing the knowledge and skills of participants and improving their dietary intake in the direction of the Mediterranean diet pattern. Moreover, our study suggests that remote health promotion interventions could offer a cost-effective community approach to address the increasing health burden.

Acknowledgments

Design and concept: LG, MB, MM, JA, and MR. Drafting of the paper: LG and MR. Analysis and interpretation of the data: LG, VD, MM, and MR. Critical revision of the paper for important intellectual content: LG, VD, MB, JA, MM, MR. Collection and assembly of data: LG, VD, MB, PR, LT, JL, EC, AC, RR, JI, JA. Final approval of the paper: LG, VD, MB, PR, LT, JL, EC, AC, RR, JI, JA, MM, MR. This work was funded by the Spanish Government Official Agency for funding biomedical research—Instituto de Salud Carlos III (ISCIII), with competitive grants through the Fondo de Investigación Sanitaria y Fondo Europeo de Desarrollo Regional (PI17/00718, PI17/00748, PI17/01870), the Regional Government of Navarra (46/2016), and the Spanish Society of Cardiology (FEC/2016). Innoliva is providing the necessary amounts of EVOO, and the Basque Culinary Center has collaborated in the elaboration of videos for the nutritional intervention. The authors thank all the volunteers of the PREDIMAR study as well as the dietitians Estíbaliz Goñi and María José Cobo, the medical doctor Liz Ruiz, and the nurses and research coordinators from each recruitment center.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clinical follow-up visits and intervention period.
[PDF File (Adobe PDF File), 97 KB - jmir_v22i12e21436_app1.pdf]

Multimedia Appendix 2
English version of the 14-item Mediterranean adherence screener (MEDAS) questionnaire.
[PDF File (Adobe PDF File), 113 KB - jmir_v22i12e21436_app2.pdf]

Multimedia Appendix 3
Screenshots of the prevention of recurrent arrythmias with Mediterranean diet (PREvención con DIeta Mediterránea de Arritmias Recurrentes) website.
[PDF File (Adobe PDF File), 94896 KB - jmir_v22i12e21436_app3.pdf]

Multimedia Appendix 4
Screenshots of the prevention of recurrent arrythmias with Mediterranean diet app.
[PDF File (Adobe PDF File), 58493 KB - jmir_v22i12e21436_app4.pdf]

Multimedia Appendix 5
Food frequency questionnaire.
[PDF File (Adobe PDF File), 4016 KB - jmir_v22i12e21436_app5.pdf]

Multimedia Appendix 6
Adherence to the Mediterranean diet at baseline, after 12- and 24-month follow-up visits according to intervention groups.
[PDF File (Adobe PDF File), 76 KB - jmir_v22i12e21436_app6.pdf]

Multimedia Appendix 7
CONSORT-eHEALTH (V 1.6.1).
[PDF File (Adobe PDF File), 268 KB - jmir_v22i12e21436_app7.pdf]

References

1. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide epidemiology of atrial fibrillation: a global burden of disease 2010 study. Circulation 2014 Feb 25;129(8):837-847 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.113.005119] [Medline: 24345399]
2. Zoni-Berisso M, Lercari F, Carazza T, Domenicucci S. Epidemiology of atrial fibrillation: European perspective. Clin Epidemiol 2014;6:213-220 [FREE Full text] [doi: 10.2147/CLEP.S47385] [Medline: 24966695]
3. Colilla S, Crow A, Petkun W, Singer DE, Simon T, Liu X. Estimates of current and future incidence and prevalence of atrial fibrillation in the US adult population. Am J Cardiol 2013 Oct 15;112(8):1142-1147. [doi: 10.1016/j.amjcard.2013.05.063] [Medline: 23831166]
4. Patel NJ, Deshmukh A, Pant S, Singh V, Patel N, Arora S, et al. Contemporary trends of hospitalization for atrial fibrillation in the United States, 2000 through 2010: implications for healthcare planning. Circulation 2014 Jun 10;129(23):2371-2379. [doi: 10.1161/CIRCULATIONAHA.114.008201] [Medline: 24842943]
5. Asad ZU, Yousif A, Khan MS, Al-Khatib SM, Stavrakis S. Catheter ablation versus medical therapy for atrial fibrillation: a systematic review and meta-analysis of randomized controlled trials. Circ Arrhythm Electrophysiol 2019 Sep;12(9):e007414. [doi: 10.1161/CIRCEP.119.007414] [Medline: 31431051]
6. Calkins H, Hindricks G, Cappato R, Kim Y, Saad EB, Aguinaga L, et al. 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. Heart Rhythm 2017 Oct;14(10):e275-e444 [FREE Full text] [doi: 10.1016/j.hrthm.2017.05.012] [Medline: 28506916]
7. Al-Hijji MA, Deshmukh AJ, Yao X, Mwangi R, Sangaralingham LR, Friedman PA, et al. Trends and predictors of repeat catheter ablation for atrial fibrillation. Am Heart J 2016 Jan;171(1):48-55. [doi: 10.1016/j.ahj.2015.10.015] [Medline: 26699600]
8. Gorenek B, Pelliccia A, Benjamin EJ, Boriani G, Crijns HJ, Fogel RI, et al. European heart rhythm association (EHRA)/European association of cardiovascular prevention and rehabilitation (EACPR) position paper on how to prevent atrial fibrillation endorsed by the heart rhythm society (HRS) and Asia pacific heart rhythm society (APHRS). Europace 2017 Feb 1;19(2):190-225 [FREE Full text] [doi: 10.1093/europace/euw242] [Medline: 28175283]
9. Martínez-González MA, Gea A, Ruiz-Canela M. The Mediterranean Diet and Cardiovascular Health. Circ Res 2019 Mar;124(5):779-798. [doi: 10.1161/CIRCRESAHA.118.313348] [Medline: 30817261]
10. Martínez-González M, Toledo E, Arós F, Fiol M, Corella D, Salas-Salvadó J, PREDIMED Investigators. Extravirgin olive oil consumption reduces risk of atrial fibrillation: the PREDIMED (Prevención con Dieta Mediterránea) trial. Circulation 2014 Jul 01;130(1):18-26. [doi: 10.1161/CIRCULATIONAHA.113.006921] [Medline: 24787471]
11. Rodrigo R, Korantzopoulos P, Cereceda M, Asenjo R, Zamorano J, Villalabeitia E, et al. A randomized controlled trial to prevent post-operative atrial fibrillation by antioxidant reinforcement. J Am Coll Cardiol 2013 Oct 15;62(16):1457-1465 [FREE Full text] [doi: 10.1016/j.jacc.2013.07.014] [Medline: 23916928]
12. Mattioli AV, Miloro C, Pennella S, Pedrazzi P, Farinetti A. Adherence to Mediterranean diet and intake of antioxidants influence spontaneous conversion of atrial fibrillation. Nutr Metab Cardiovasc Dis 2013 Feb;23(2):115-121. [doi: 10.1016/j.numecd.2011.03.005] [Medline: 21798731]
13. Bazal P, Gea A, de la Fuente-Arrillaga C, Barrio-López MT, Martinez-González MA, Ruiz-Canela M. Olive oil intake and risk of atrial fibrillation in the SUN cohort. Nutr Metab Cardiovasc Dis 2019 May;29(5):450-457. [doi: 10.1016/j.numecd.2019.02.002] [Medline: 30948307]
14. Bazal P, Gea A, Martínez-González MA, Salas-Salvadó J, Asensio EM, Muñoz-Bravo C, et al. Mediterranean alcohol-drinking pattern, low to moderate alcohol intake and risk of atrial fibrillation in the PREDIMED study. Nutr Metab Cardiovasc Dis 2019 Jul;29(7):676-683. [doi: 10.1016/j.numecd.2019.03.007] [Medline: 31078364]
15. Larsson SC, Wolk A. Fish, long-chain omega-3 polyunsaturated fatty acid intake and incidence of atrial fibrillation: A pooled analysis of two prospective studies. Clin Nutr 2017 Apr;36(2):537-541. [doi: 10.1016/j.clnu.2016.01.019] [Medline: 26875446]
16. Larsson SC, Drca N, Björck M, Bäck M, Wolk A. Nut consumption and incidence of seven cardiovascular diseases. Heart 2018 Oct;104(19):1615-1620 [FREE Full text] [doi: 10.1136/heartjnl-2017-312819] [Medline: 29661934]
17. Bazal P, Gea A, Navarro AM, Salas-Salvadó J, Corella D, Alonso-Gómez A, et al. Caffeinated coffee consumption and risk of atrial fibrillation in two Spanish cohorts. Eur J Prev Cardiol 2020 Mar 4:2047487320909065. [doi: 10.1177/2047487320909065] [Medline: 32131628]
18. Larsson SC, Drca N, Jensen-Urstad M, Wolk A. Chocolate consumption and risk of atrial fibrillation: two cohort studies and a meta-analysis. Am Heart J 2018 Jan;195:86-90. [doi: 10.1016/j.ahj.2017.09.013] [Medline: 29224650]
19. Gronroos NN, Alonso A. Diet and risk of atrial fibrillation – epidemiologic and clinical evidence –. Circ J 2010 Oct;74(10):2029-2038 [FREE Full text] [doi: 10.1253/circj.cj-10-0820] [Medline: 20838006]
20. Barrio-Lopez MT, Ruiz-Canela M, Ramos P, Tercedor L, Ibañez Criado JL, Ortiz M, et al. PREvention of recurrent arrhythmias with Mediterranean diet (PREDIMAR) study in patients with atrial fibrillation: Rationale, design and methods. Am Heart J 2020 Feb;220:127-136. [doi: 10.1016/j.ahj.2019.10.009] [Medline: 31809992]
21. Willis EA, Szabo-Reed AN, Ptomey LT, Steger FL, Honas JJ, Washburn RA, et al. Do weight management interventions delivered by online social networks effectively improve body weight, body composition, and chronic disease risk factors?

A systematic review. J Telemed Telecare 2017 Feb;23(2):263-272 [FREE Full text] [doi: 10.1177/1357633X16630846] [Medline: 26880695]

22. Afshin A, Babalola D, Mclean M, Yu Z, Ma W, Chen C, et al. Information technology and lifestyle: a systematic evaluation of internet and mobile interventions for improving diet, physical activity, obesity, tobacco, and alcohol use. J Am Heart Assoc 2016 Aug 31;5(9):- [FREE Full text] [doi: 10.1161/JAHA.115.003058] [Medline: 27581172]
23. Teasdale N, Elhussein A, Butcher F, Piernas C, Cowburn G, Hartmann-Boyce J, et al. Systematic review and meta-analysis of remotely delivered interventions using self-monitoring or tailored feedback to change dietary behavior. Am J Clin Nutr 2018 Feb 1;107(2):247-256 [FREE Full text] [doi: 10.1093/ajcn/nqx048] [Medline: 29529158]
24. Contento IR. Nutrition education: linking research, theory, and practice. Asia Pac J Clin Nutr 2008;17(Suppl 1):176-179 [FREE Full text] [Medline: 18296331]
25. Schröder H, Fitó M, Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr 2011 Jun;141(6):1140-1145. [doi: 10.3945/jn.110.135566] [Medline: 21508208]
26. Blake CE, Fisher JO, Ganter C, Younginer N, Orloski A, Blaine RE, et al. A qualitative study of parents' perceptions and use of portion size strategies for preschool children's snacks. Appetite 2015 May;88:17-23 [FREE Full text] [doi: 10.1016/j.appet.2014.11.005] [Medline: 25447008]
27. Gibson AA, Hsu MS, Rangan AM, Seimon RV, Lee CM, Das A, et al. Accuracy of hands vs household measures as portion size estimation aids. J Nutr Sci 2016;5:e29 [FREE Full text] [doi: 10.1017/jns.2016.22] [Medline: 27547392]
28. Sánchez-Tainta A, San Julián B, Martínez-González MA. PREDIMED date el gusto de comer sano. Barañain: EUNSA; 2015.
29. Martínez-González M. Salud a ciencia cierta. Consejos para una vida sana (sin caer en las trampas de la industria). Barcelona: Planeta; 2018.
30. Martin-Moreno JM, Boyle P, Gorgojo L, Maisonneuve P, Fernandez-Rodriguez JC, Salvini S, et al. Development and validation of a food frequency questionnaire in Spain. Int J Epidemiol 1993 Jun;22(3):512-519. [doi: 10.1093/ije/22.3.512] [Medline: 8359969]
31. Martínez-González MA, López-Fontana C, Varo JJ, Sánchez-Villegas A, Martinez JA. Validation of the Spanish version of the physical activity questionnaire used in the Nurses' Health Study and the Health Professionals' Follow-up Study. Public Health Nutr 2005 Oct;8(7):920-927. [doi: 10.1079/phn2005745] [Medline: 16277809]
32. Crichton GE, Howe PR, Buckley JD, Coates AM, Murphy KJ, Bryan J. Long-term dietary intervention trials: critical issues and challenges. Trials 2012 Jul 20;13:111 [FREE Full text] [doi: 10.1186/1745-6215-13-111] [Medline: 22812577]
33. Moreiras O, Carbajal A, Cabrera L, Cuadrado C. Tablas de composición de alimentos: Guía de prácticas (Spanish food composition tables). 7th edition. Madrid: Ediciones PIRAMIDE; 2003.
34. Mataix J, García L, Mañas M, Martínez DV, Llopis J. Tabla de composición de alimentos (Spanish food composition tables). 4th edition. Granada: Universidad de Granada; 2003.
35. Encuesta Enalia 2. Encuesta Nacional De Alimentación en Población Adulta, Mayores Y Embarazadas. Agencia Española de Seguridad Alimentaria y Nutrición (AESAN). URL: http://www.aecosan.msssi.gob.es/AECOSAN/web/seguridad_alimentaria/subdetalle/enalia_2.htm [accessed 2020-09-10]
36. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018 Apr 11;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]
37. Hutchesson MJ, Collins CE, Morgan PJ, Watson JF, Guest M, Callister R. Changes to dietary intake during a 12-week commercial web-based weight loss program: a randomized controlled trial. Eur J Clin Nutr 2014 Jan;68(1):64-70. [doi: 10.1038/ejcn.2013.194] [Medline: 24129359]
38. Duncan M, Vandelanotte C, Kolt GS, Rosenkranz RR, Caperchione CM, George ES, et al. Effectiveness of a web- and mobile phone-based intervention to promote physical activity and healthy eating in middle-aged males: randomized controlled trial of the ManUp study. J Med Internet Res 2014 Jun 12;16(6):e136 [FREE Full text] [doi: 10.2196/jmir.3107] [Medline: 24927299]
39. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. J Med Internet Res 2010 Dec 14;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]
40. Zazpe I, Sanchez-Tainta A, Estruch R, Lamuela-Raventos RM, Schröder H, Salas-Salvado J, et al. A large randomized individual and group intervention conducted by registered dietitians increased adherence to Mediterranean-type diets: the PREDIMED study. J Am Diet Assoc 2008 Jul;108(7):1134-44; discussion 1145. [doi: 10.1016/j.jada.2008.04.011] [Medline: 18589019]
41. Estruch R, Ros E, Salas-Salvadó J, Covas M, Corella D, Arós F, PREDIMED Study Investigators. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. N Engl J Med 2018 Jun 21;378(25):e34. [doi: 10.1056/NEJMoa1800389] [Medline: 29897866]
42. Sayón-Orea C, Razquin C, Bulló M, Corella D, Fitó M, Romaguera D, et al. Effect of a nutritional and behavioral intervention on energy-reduced mediterranean diet adherence among patients with metabolic syndrome: interim analysis of the

predimed-plus randomized clinical trial. J Am Med Assoc 2019 Oct 15;322(15):1486-1499 [FREE Full text] [doi: 10.1001/jama.2019.14630] [Medline: 31613346]

43. Quintana-Navarro GM, Alcala-Diaz JF, Lopez-Moreno J, Perez-Corral I, Leon-Acuña A, Torres-Peña JD, et al. Long-term dietary adherence and changes in dietary intake in coronary patients after intervention with a Mediterranean diet or a low-fat diet: the CORDIOPREV randomized trial. Eur J Nutr 2020 Aug;59(5):2099-2110. [doi: 10.1007/s00394-019-02059-5] [Medline: 31342228]
44. Recio-Rodriguez JI, Agudo-Conde C, Martin-Cantera C, González-Viejo MN, Fernandez-Alonso MD, Arietaleanizbeaskoa MS, EVIDENT Investigators. Short-term effectiveness of a mobile phone app for increasing physical activity and adherence to the Mediterranean diet in primary care: a randomized controlled trial (evident II study). J Med Internet Res 2016 Dec 19;18(12):e331 [FREE Full text] [doi: 10.2196/jmir.6814] [Medline: 27993759]
45. Choi BG, Dhawan T, Metzger K, Marshall L, Akbar A, Jain T, et al. Image-based mobile system for dietary management in an american cardiology population: pilot randomized controlled trial to assess the efficacy of dietary coaching delivered via a smartphone app versus traditional counseling. JMIR Mhealth Uhealth 2019 Apr 23;7(4):e10755 [FREE Full text] [doi: 10.2196/10755] [Medline: 31012860]
46. van Dijk MR, Koster MP, Oostingh EC, Willemsen SP, Steegers EA, Steegers-Theunissen RP. A mobile app lifestyle intervention to improve healthy nutrition in women before and during early pregnancy: single-center randomized controlled trial. J Med Internet Res 2020 May 15;22(5):e15773 [FREE Full text] [doi: 10.2196/15773] [Medline: 32412417]
47. Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, et al. A fully automated web-based program improves lifestyle habits and hba1c in patients with type 2 diabetes and abdominal obesity: randomized trial of patient e-coaching nutritional support (the anode study). J Med Internet Res 2017 Nov 8;19(11):e360 [FREE Full text] [doi: 10.2196/jmir.7947] [Medline: 29117929]
48. Springvloet L, Lechner L, de Vries H, Oenema A. Long-term efficacy of a web-based computer-tailored nutrition education intervention for adults including cognitive and environmental feedback: a randomized controlled trial. BMC Public Health 2015 Apr 12;15:372 [FREE Full text] [doi: 10.1186/s12889-015-1707-4] [Medline: 25887891]
49. Lindström J, Peltonen M, Eriksson JG, Ilanne-Parikka P, Aunola S, Keinänen-Kiukaanniemi S, Finnish Diabetes Prevention Study (DPS). Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). Diabetologia 2013 Feb;56(2):284-293. [doi: 10.1007/s00125-012-2752-5] [Medline: 23093136]
50. Shai I, Schwarzfuchs D, Henkin Y, Shahar DR, Witkow S, Greenberg I, Dietary Intervention Randomized Controlled Trial (DIRECT) Group. Weight loss with a low-carbohydrate, Mediterranean, or low-fat diet. N Engl J Med 2008 Jul 17;359(3):229-241. [doi: 10.1056/NEJMoa0708681] [Medline: 18635428]
51. Guasch-Ferré M, Liu G, Li Y, Sampson L, Manson JE, Salas-Salvadó J, et al. Olive oil consumption and cardiovascular risk in US adults. J Am Coll Cardiol 2020 Apr 21;75(15):1729-1739. [doi: 10.1016/j.jacc.2020.02.036] [Medline: 32147453]
52. Consumo Alimentario Anual en Los Hogares Españoles 2019. Ministerio de Agricultura Pesca y Alimentación. URL: https://www.mapa.gob.es/es/alimentacion/temas/consumo-tendencias/panel-de-consumo-alimentario/ultimos-datos/default.aspx [accessed 2020-09-15]
53. USDA Food Composition Database. US Department of Agriculture. URL: https://ndb.nal.usda.gov [accessed 2020-10-08]
54. Neveu V, Perez-Jiménez J, Vos F, Crespy V, du Chaffaut L, Mennen L, et al. Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. Database (Oxford) 2010;2010:bap024 [FREE Full text] [doi: 10.1093/database/bap024] [Medline: 20428313]

Abbreviations

AF: atrial fibrillation
EVOO: extra virgin olive oil
FFQ: food frequency questionnaire
MEDAS: Mediterranean diet adherence screener
MUFAs: monounsaturated fatty acids
PREDIMAR: prevention of recurrent arrythmias with Mediterranean diet (PREvención con DIeta Mediterránea de Arritmias Recurrentes)
PREDIMED: prevention with Mediterranean diet (PREvención con DIeta MEDiterránea)
PUFAs: polyunsaturated fatty acids
SFA: saturated fatty acids
VOO: virgin olive oil

Edited by G Eysenbach; submitted 16.06.20; peer-reviewed by S Santarossa, X Li; comments to author 17.08.20; revised version received 12.10.20; accepted 28.10.20; published 07.12.20.

Please cite as:
Goni L, de la O V, Barrio-López MT, Ramos P, Tercedor L, Ibañez-Criado JL, Castellanos E, Ibañez Criado A, Macias Ruiz R, García-Bolao I, Almendral J, Martínez-González MÁ, Ruiz-Canela M
A Remote Nutritional Intervention to Change the Dietary Habits of Patients Undergoing Ablation of Atrial Fibrillation: Randomized Controlled Trial
J Med Internet Res 2020;22(12):e21436
URL: http://www.jmir.org/2020/12/e21436/
doi:10.2196/21436
PMID:33284131

Original Paper

Digital Health Interventions for People With Type 2 Diabetes to Develop Self-Care Expertise, Adapt to Identity Changes, and Influence Other's Perception: Qualitative Study

Sophie Turnbull[1], BSc, MSc, PhD; Patricia J Lucas[2], BSc, PhD; Alastair D Hay[1], MBChB, MD (Leic), MRCP, DCH (RCP), MRCGP, DFFP, FHEA; Christie Cabral[1], BSc, MRes, PhD

[1]Centre for Academic Primary Care, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[2]School for Policy Studies, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Sophie Turnbull, BSc, MSc, PhD
Centre for Academic Primary Care
Population Health Sciences
Bristol Medical School, University of Bristol
Canynge Hall
39 Whatley Road
Bristol, BS8 2PS
United Kingdom
Phone: 44 7855 115
Email: sophie.turnbull@bristol.ac.uk

Abstract

Background: A diagnosis of type 2 diabetes (T2D) results in widespread changes to a person's life and can be experienced as an assault on their sense of self. The resources available to an individual influence how the individual adapts to their diabetic identity and subsequently engages in self-care. Digital interventions can be viewed as a resource that people can draw on to adapt to the diagnosis. However, there is an indication that people from disadvantaged groups find digital health technologies more challenging to access and use, which may increase health inequalities.

Objective: This study aims to gain insights into how and why people with T2D use digital self-care technology and how experiences vary between individuals and social groups.

Methods: A purposive sample of people who had used a digital intervention to help them self-care for their T2D were recruited for the study. Semistructured interviews were conducted, and data were analyzed thematically.

Results: A diverse sample of 21 participants were interviewed. Participants used digital interventions to help them to understand and feel more in control of their bodies. Digital interventions were used by participants to project their chosen identity to others. Participants selected technology that allowed them to confirm and enact their preferred positive identities, both by avoiding stigma and by becoming experts in their disease or treatment. Participants preferred using digital interventions that helped them conceal their diabetes, including by buying discrete blood glucose monitors. Some participants used technology to increase their sense of power in their interaction with clinicians, whereas others used technology to demonstrate their *good*ness.

Conclusions: The technology that people with T2D have access to shapes the way they are able to understand and control their bodies and support preferred social identities.

(J Med Internet Res 2020;22(12):e21328) doi:10.2196/21328

KEYWORDS

social identity; diabetes mellitus; internet-based intervention; health care disparities; socioeconomic factors; self care

Introduction

Type 2 diabetes (T2D) is a chronic disease that affects a large number of people and creates a significant burden for patients and the health services that support them [1-5]. Worldwide, there is a prevalence rate of 6059 per 100,000 people, and the prevalence continues to rise [6]. In the United Kingdom, 1 in 10 people over the age of 40 years now has T2D [6]. Self-care is the dominant medical paradigm for managing chronic illnesses in wealthy nations, such as the United Kingdom [7,8]. The self-care model holds the individual responsible for managing their chronic illness [9]. For example, in the case of T2D, people are asked to change their diet; lose weight; administer their own medications; and in some cases, monitor their own blood glucose levels [5,10]. The self-care model presumes that an individual has both power and accountability to change the factors that affect their health and expects the person to take control of their condition [11,12]. The focus of individual responsibility in the self-care model neglects the real constraints on people's lives and the impact of social and economic circumstances on people's choices [12-13].

The self-care model demands that people consider themselves to be ill and need to change as a result of their illness. Greater engagement with management of diabetes has been shown to be associated with incorporating disease diagnosis into one's social identity [13,14]. It is argued that becoming chronically ill and the associated changes to lifestyle are a challenge to self-identity [14,15]. A new social identity must be negotiated and incorporated into their existing social identities [16]. If successful, the illness is experienced as *just a part* of them [16,17]. For those who are unable to reconcile the illness identity with other social identities, the diagnosis may present a sense of the *diminished self* [17,18]. Understanding how people incorporate illness identities is, therefore, crucial to understanding self-care.

However, another challenge of the self-care model is that it assumes that everyone has the resources and capacity to self-manage [19-21] and that managing oneself is desirable. In terms of the former, there is established literature demonstrating that people in more privileged social positions have greater access to resources that can be used to avoid risk and to adopt protective strategies [15,19-24]. Health literacy is a prime example of this; where it is built on a foundation of early advantage or disadvantage and mediates any benefit from health education [25]. Furthermore, the promotion of self-control as a desirable expression of agency and power is culturally specific [18]. This Eurocentric interpretation sees loss of control as indicative of chaos and moral failing [11,13,15,18]. Within this individualistic framing, diagnosis is understood as a loss of control. and the inability to manage the disease as an individual's moral failure [15]. Those who struggle to manage their illness are stigmatized and may receive less support from professionals, friends, family, and other people with diabetes [11,15]. This means that some groups are (likely) served poorly by traditional education campaigns that promote self-care: those with fewer financial and social resources to enable change, those with limited health literacy, those who struggle to assimilate their illness into their social identity, and those for whom self-control is not an inevitable and moral good.

Digital health interventions are an increasingly common tool used for self-care for people living with T2D [26]. As Lupton [27] puts it "digital health technologies are positioned as contributing to lay people effectively becoming the 'managers' of their own health and health care." One argument for this shift is that digital health may reduce inequalities by increasing access to health interventions and by providing interventions that are tailored to the needs of the user [28-30]. The literature on how digital technology may be useful in supporting identity shifts is beginning to build [31], and one study that comments on the identity work that happens in online diabetes support groups [32]. This qualitative interview study was conducted to explore the experiences of adults with T2D using digital self-care technology. Our aim is to explore how adult users talk about their use of digital interventions for self-management of T2D by examining how they spoke about their identity in relation to their technology use and their illness.

Methods

Overview

The methodological orientation used in this study was an inductive approach drawing on the aspects of grounded theory [33,34]. Data were gathered using semistructured one-to-one interviews.

Ethical approval was granted by the University of Bristol, Faculty of Health Sciences Research Ethics Committee (the approval letter is provided in Multimedia Appendix 1).

Participants

Participants were approached face to face and by email through community and diabetes groups, with a focus on groups that served lower-income neighborhoods and Black, Asian, and Minority Ethnic groups. Adverts were also placed in the *Diabetes UK* web-based and print magazines. We sought out adults who had been diagnosed with T2D and had used a digital intervention to support them to manage their condition at least once and who spoke and understood English. Recruitment materials sought out a range of experiences with digital interventions, including those who did not like them. The recruitment materials asked anyone interested the study to complete a screening questionnaire (Multimedia Appendix 2) that enabled purposive sampling of participants to capture a range of experiences across different social groups. A total of 27 potential participants completed the screening questionnaire, and interviews continued until data saturation was reached for the major themes.

Procedure

Theory and previous research on the lived experience of T2D, self-care, and the digital divide were used to develop the topic guide. The topic guide was also developed iteratively, with revisions made to reflect themes emerging from the analysis. The 3 iterations of the topic guide are available in Multimedia Appendix 3.

Interviews were conducted in person or by telephone, according to the preferences of the participants. In-person interviews were conducted in Bristol and Leicester, and participants could choose to be interviewed in their homes or in a hospital-based diabetes unit. Written and verbal information about the research was provided, and the participants provided consent ahead of taking part in the interview. Interview duration ranged from 35 minutes to 2 hours and 13 minutes, and they were audio recorded and transcribed verbatim. All the interviews were conducted by 1 researcher (ST). ST took field notes during and immediately after the interviews.

Analysis

Analysis began soon after data collection was started and was ongoing and iterative. Interviewing continued until data saturation was reached, and no new data arose in relation to the key themes. Encrypted audio recorders were used to record the interviews. The audio was transcribed verbatim and anonymized, and the script was checked for accuracy and imported into NVivo 12 (qualitative data analysis software; QSR International) for analysis. Both the audio and the transcripts were stored on the secure servers at the University of Bristol and in accordance with the Data Protection Act 2018. A thematic approach was used to analyze the data [35]. Theoretical frameworks and evidence from studies exploring the lived experience of chronic illness, self-care and health inequalities was used to identify some major themes before coding, and further themes emerged from the data. The theoretical frameworks and evidence included chronic illness as an assault on personal identities; stigma, self-determination, and control in chronic illness; the moral component of self-care; and the influence of the socioeconomic context on how people adapt to a chronic illness and their ability to self-care for their condition. The first 3 transcripts were independently coded by 3 authors (ST, CC, and PL). This team reviewed the coded transcripts in a meeting and developed a coding structure by consensus. Regular meetings were held to review subsequent coding, and the coding structure was adapted to accommodate new themes. The full coding tree is available in Multimedia Appendix 4. Participants were provided with a summary of the main findings after the analysis was complete.

Research Team and Reflexivity

Personal Characteristics

This study was conducted as a component of ST's PhD, during which she received formal and informal training in qualitative methods and was supervised by CC and PL, who are senior academics specializing in qualitative research. ST's previous qualifications were a BSc degree in psychology and an MSc degree in neuropsychology, and most of her training and experience was in quantitative methods, which may have had a bearing on the conduct and the interpretation of the interviews. CC has a background in anthropology, and PL has a background in psychology; however, both work on applied health topics now, with a focus on inequalities in health.

Relationship With Participants

The study participants were unknown to the authors before the commencement of this study. The participants would have known that ST was a researcher at the University of Bristol. The participants who were interviewed in person would be aware that the interviewer (ST) was a White woman in her 30s, who appears to be relatively affluent, at a healthy weight, and with no visible disabilities. The participants were aware that the aim of the study is to explore the use of technology to support the self-management of T2D; however, they would not have known that the authors were exploring accessibility and identity. ST's position while conducting the interviews was that digital interventions have the potential to be beneficial for people with chronic conditions and that there are likely to be differences in the way people access and use technology, particularly reflecting social and cultural differences.

Results

Sample Description

A total of 21 people were interviewed. One person initially expressed an interest in the study and then refused participation because they were uncomfortable with the university standard procedures for the storage of interview data. The brief sample overview is shown in Table 1 and the individual participant profile is shown in Multimedia Appendix 5. The sample achieved diversity in gender, household income, and neighborhood deprivation. Most participants were White British (White British, n=17; Asian or British Asian, n=3; other White background, n=1) and Christian, and the majority of our sample were older and well-educated individuals. Two-thirds of the participants had a university degree or equivalent, and none of the participants reported leaving school before the age of 16 years. There were 6 potential participants who were not interviewed because their characteristics were similar to the majority of the sample (White British, well educated, and from more affluent neighborhoods) and, therefore, did not increase the diversity of the purposive sample.

Table 1. Participant characteristics (N=21).

Participant characteristics	Male (n=11), n (%)	Female (n=10), n (%)
Education		
Secondary school or equivalent (low education)	1 (9)	0 (0)
Intermediate between secondary level and university (eg, NVQ3-5[a], diploma, and apprenticeship; low education)	5 (46)	2 (20)
University degree or equivalent (high education)	5 (46)	8 (80)
Estimated household income in the last year (before tax and not including benefits), £ (US $)		
Lowest income: <16,000 (<21,332) or eligible for means tested benefits	3 (27)	0 (0)
Low income: 16,000-24,999 (21,332-33,330)	1 (9)	4 (40)
Mid income: 25,000-34,999 (33,331-46,662)	3 (27)	0 (0)
High income: 35,000-44,999 (46,663-59,994)	0 (0)	2 (20)
Highest income: >45,000 (>59,995)	2 (18)	2 (20)
Prefer not to say	2 (18)	2 (20)
Use of digital interventions		
Lighter (≤2 interventions)	7 (64)	5 (50)
Heavier (>2 interventions)	4 (36)	5 (50)
Home neighborhood deprivation[b]		
1 Most deprived	1 (9)	2 (20)
2 Lower SES[c]	2 (18)	1 (10)
3 Mid SES	3 (27)	1 (10)
4 Higher SES	1 (9)	2 (20)
5 Highest SES	4 (36)	3 (30)
Not available	0 (0)	1 (10)
Age (years)		
21-40	1 (9)	1 (10)
41-60	4 (36)	5 (50)
61-70	6 (55)	4 (40)
71-80	2 (18)	1 (10)

[a] NVQ3-5: National Vocational Qualification levels 3 to 5.

[b] Indices of multiple deprivation score derived from the participant's home post code were used to determine the participant's neighbourhood deprivation within the United Kingdom, and the quintile is given.

[c] SES: socioeconomic status.

The sample was evenly divided between light and heavy users of digital health interventions. Most participants did not use interventions designed specifically for people with diabetes but rather used technology designed to support healthy living and social connectivity. Wearable fitness trackers were the most commonly used intervention (16 participants) and apps that tracked nutrition or fitness (11 participants). The diabetes-specific interventions were the blood glucose monitors (BGMs; Dario meter, Freestyle Libre, and Trueyou mini) used by 10 participants (5 supplied by health care practitioners [HCPs] and 5 purchased privately) and 3 different apps each used by 1 participant (Diabetes diary, IBG star app, and Habits, a South Asian–specific diabetes app). The median number of different interventions trialed by the participants was 2 (range 1-7); 12 participants were lighter users (≤2 intervention) and 9 were heavier users (>2 interventions) of digital technology.

Findings

As our focus is on how participants interacted with digital health interventions, we focused on the ways in which they were used regardless of whether they were light or heavy users of this technology. Most participants (both heavier and lighter users) used fitness trackers. Users who had tried a greater number and range of different types of technology were heavier users, whereas the 3 users who used diabetes-specific self-care apps (excluding BGMs) were lighter users.

Participants used technology to help them understand their body and to feel like they had more control over their bodies and their diabetes. Digital interventions were used by participants to resist stigmatizing illness identities and to project a positive identity. Participants also used technology to increase their sense of power or status in their interaction with HCPs.

Understanding Their Bodies and Making the Invisible, Visible

Digital interventions were used as a tool to help participants understand their bodies, to develop their expertise in self-care, and to keep them engaged in the long-term management of their condition. Participants talked about using feedback from BGMs to monitor their bodies and establish how their "*body works, how it reacts*" [Participant number 33], turning a relatively hidden illness into something visible and tangible:

> *I put myself to test my own blood, finger prick testing...because you have nowhere to hide from that evidence.* [Participant number 33, female, high education, heavier user]

One woman described herself as an "*inveterate self-experimenter*" [Participant number 41, female, high education, heavier user], describing trial-and-error experiments, in which she tested whether specific diets recommended on the web worked for her, using the output from monitors as evidence.

Similarly, feedback from wearable fitness monitors was used by participants to see changes in their own fitness or behaviors:

> *The heart rate monitor, it's really good because...you can get out of breath but your heart rate...can come down very quickly. You realise that your fitness levels are going up....I felt really excited by it; liberated actually.* [Participant number 37, female, high education, heavier user]

> *I wouldn't be without my Fitbit, it drives me. It absolutely drives me, because I get panicky last week when I couldn't blooming recharge the thing...I have to know what I'm doing.* [Participant number 10, female, low education, lighter user]

Wearable technology supported their self-care by providing motivation to be more active and positive feedback on their achievements:

> *The monitoring [with a Fitbit] gives me a reward. (...)an aid to help me celebrate my achievement.* [Participant number 30, female, high education, heavier user]

Feeling in Control

Through the use of digital interventions, participants felt like they had more control over their diabetes and felt more in control generally. Being "*more informed*" [Participant number 26, male, low ed] about diabetes in general and having personalized information created a feeling of greater agency to affect their diabetic bodies, behavior, and health care:

> *I think the Garmin and the, erm, er, it, it is good really...the more you start seeing what your body is doing, the more fascinating it becomes, and you feel more in control, really.*" [Participant number 37, female, high education, heavier user]

One woman spoke about how having web-based access to her medical records made her feel like she had more ownership over her health care, feeling like she was "*part of it, and it's not something the doctor owns*," which made her "*a bit more focused to try and get*" her blood glucose levels "*better, get more under control*" [Participant number 31, female, low education, light user].

Digital interventions helped participants to feel more in control of their diabetes in situations where they were out of their normal routine. This included when people were in environments where they could not control what happens, such as holidays, and when they had changed their management strategies:

> *I- mainly [use the Freestyle Libre] when I'm at most risk of going off, off, erm, the wagon. (...) if I have any work trips and I'm staying in a hotel, I'll slap one on, because that way, as I say, it gives me more self-control(...) without those two things [blood glucose meter and Freestyle Libre], I wouldn't be in control of my blood glucose. I would, I would be thinking, "Oh well, just one won't hurt, will it? This is a special dinner, I'll have pudding."* [Participant number 41, female, high education, heavier user]

Projecting Positive Identities

Participants used digital interventions to project, enact, and confirm their preferred positive identities. Through their use of technology, participants presented themselves as someone who "*got things under control*" [Participant number 27, male, low education, light user]. Participants often described the use of digital technologies as being associated with being younger, fitter, in control, and more skilled or educated; with higher status; or with specialized knowledge:

> *It's probably wrong to try and box people in, but I don't think it's any point trying to tell an 85-year-old about Fitbits.* [Participant number 27, male, low education, light user]

> *...we've always had Fitbits....from when they were first introduced into the UK...we've just upgraded it to a newer model...some mornings I'll get up and...I'll just wear my Apple watch for the day just for something different.* [Participant number 34, male, low education, light user]

Some described how the technology they used required a level of understanding that not everyone had. One man talked about how he could understand the information he got from the Freestyle Libre because he had done a "*mathematics Open University, er, foundation course*" and consequently could "*understand a bit about the statistics*" [Participant number 42, male, high education, heavier user]. A woman described how she would look at the data from her Freestyle Libre each morning because she was a "*data master(...) someone who likes data*" [Participant number 33, female, high education, heavier user]. These were reflections of personal attributes rather than demographic differences in this sample; younger participants [Participant numbers 23 and 31] were lighter users, whereas others recognized that older people could acquire the right skills and knowledge:

> *...[for] the older generation apps it's not really a thing...unless you're, silver surfers....they have to understand what the app is doing or what is an app, 'cause a lot of people don't understand what an app is.* [Participant number 38, male, higher education, heavier user]

Defending Against the Stigmatized Diabetic Identity

The stigma felt by participants in relation to their diabetes diagnosis was apparent in interviews. Many of the participants described experiencing stigma as a result of being diagnosed with diabetes. There was a sense that family, friends, and the media blamed them for getting diabetes because they were overweight or (they presumed) they had an unhealthy lifestyle:

> *I remember somebody saying to me, "Well, do you think it's 'cause you probably drank too much?"* [Participant number 11, female, high education, heavier user]

> *...people think you're a druggy....I remember doing it [injecting insulin],... in a council office, I was doing consultancy work and my director at the time said, "You shouldn't be dong that here..."* [Participant number 38, male, high education, heavier user]

Some talked about being given unsolicited advice on diet and exercise from people in their social circle and HCPs:

> *...all the publicity around type 2 is entirely negative, so people think A), that I must have brought it on myself, (...) people assume that you're not exercising and they start lecturing you about that* [Participant number 30, female, high education, heavier user]

One participant talked about being stared at when she injected insulin in a public place where a man "*kept looking and looking and despite how I kind of turned or, you know, tried to move away*" [Participant number 31, female, low education, and light user].

In response to this stigma, some participants described not disclosing their diabetes diagnosis in fear of being labeled with the stigmatized diabetic identity. The 2 youngest (aged 29 and 31 years) participants who were both of British South Asian ethnicity emphasized the challenges of being diagnosed young and how that affected their wish to conceal their diagnosis because they are "*embarrassed or ashamed of it*" and because of the fear that people would "*judge you*" [Participant number 23, 31 years, male, light user]. One woman talked about not wanting to be identified as "*the diabetic lady,*" as she felt it was reductive and "*inhumane*" [Participant number 33, female, high education, heavier user]. A man spoke about people in the South Asian community concealing their diabetes diagnosis, for fear of "*their family members being tarnished,*" and emphasized that this was a problem particularly women in the community because it "*could be a barrier for her future kind of marriage proposal*" [Participant number 26, male, low education, light user]. These experiences of stigma matter particularly here because participants explicitly preferred digital health innovations, as they helped them conceal their diabetes. For example, the equipment used could aid concealment or risk exposure. One woman talked about the blood glucose testing kits supplied by the National Health Service (NHS) being "*bulky,*" and the small monitor she purchased herself allowed her to be more "*discrete*" conducting self-care activities [Participant number 40, female, low income, light user]. In contrast, a younger woman "*didn't tell anyone*" about her diagnosis; however, her condition was exposed to family members because they "*noticed*" the blood glucose "*machine,*" and she talked about how she wanted future technology to be "*a bit more discreet*" [Participant number 24, female, low income, light user]. Men also mentioned wishing to conceal, or not "*advertise*" [Participant number 22 and light user], their diagnosis, describing going to the "*gents*" or somewhere "*private*" to inject insulin [Participant number 20, male, low education, light user]. One man had stated he was happy to "*tell everyone*" about his diagnosis, but he felt strongly that the behavior of injecting insulin was "*not normal*" [Participant number 28, male, low education, light user]. Although none of the men in this sample mentioned the use of digital technology in this regard.

A rejection of a *diabetic identity* could be seen in the views of those who did not see diabetes as a progressive illness and believed that it was possible to reverse or halt their diabetes. They had not surrendered to their *sick self* and felt that their illness was something that they were still able to master. Participants described hearing stories where "*people lost weight and their diabetes actually went*" [Participant number 22, male, low education, and light user], which motivated them to go to the gym or seek out technology that could help them in their efforts "*Trying to reverse*" their diabetes [Participant number 23, 29 years, male, high education, light user], where medically they were no longer considered to be diagnosed with the condition. However, this also resulted in conflicting identities. One man was positive about technology because it had worked to reverse a diabetic diagnosis; however, he also acknowledged that he would always *have* diabetes:

> *[using apps] seems, it seems to have worked...if I went to the doctors now, I would no longer be diagnosed as diabetic. But, erm, but, that I am diagnosed as diabetic, means that, you know, in a sense, once you've got it, you know, you've got it.* [Participant number 29, male, high education, heavier user]

Increasing Status With HCPs

One of the ways that digital technology was used by participants to defend against an illness identity was as a tool to gain power in their interaction with HCPs. Some participants talked about using technology to resist treatment prescribed by clinicians or to modify their treatment regime. For example, one man bought his own blood pressure monitor to avoid taking medication:

> *obviously the GP didn't like me not taking any blood pressure tablets. Er, so I said, well, "Then, we'll keep an eye on it." "If it starts going up, I'll take the damn things."* [Participant number 36, male, high education, heavier user]

Others described how gaining knowledge of diabetes online enabled them to negotiate care, as they knew more about their condition and treatment choices:

> *you're slightly informed, so they can't just treat you as somebody who's, you know, like a naughty boy.* [Participant number 26, male, low education, light user]

One woman had received additional interest from her HCP and had been spoken to by medical students because of her weight loss, which she attributed to her Fitbit. She was treated as an expert patient, and through this, she felt a positive and desirable illness identity:

> *I said to him [GP], I braved him (...)when he asked me to see these students (...) I said to him, "If I'm doing all this for you..." Of course it's helping me, I'm enjoying telling people about my journey [losing weight with the Fitbit]. And it is successful so far...So don't get rid of me yet.* [Participant number 10, female, low education, low income, light user]

In *braving* the general practitioner, she describes both the power imbalance, her redress of this by providing a role as the expert patient through describing her journey to his students, and a demand she now felt able to make of him (that he does not get rid of her).

Some participants described using the digital interventions to provide proof of their management activities to their HCP, to demonstrate their *good*ness, and to avoid chastisement:

> *My diabetes nurse here, she's quite pleased with me. And she said (...) "Oh, I wish everybody would have one [Fitbit]"* [Participant number 35, female, high education, light user]

Discussion

Principal Findings

Participants in this study used technology to help them understand their body, develop their expertise in self-care, and keep them engaged in the management of their condition. Through the use of digital interventions, participants could better understand and control their bodies and their diabetes.

Digital interventions were used by participants to project their chosen identity to others. Participants selected technology that allowed them to confirm and enact their preferred positive identities, both by avoiding stigma and by becoming experts in their disease or treatment. Participants preferred using digital health innovations that helped them conceal their diabetes, including by buying discrete BGMs. Participants used technology to increase their power or status in their interaction with HCPs. Some participants used technology to resist treatments prescribed by clinicians, to modify their treatment regime, or to negotiate support received. Others used digital interventions to provide proof of their management activities to their HCP, to demonstrate their *goodness*, and to avoid chastisement.

Strengths and Limitations

To the authors' knowledge, this is the first study to explore why and how people choose technology to support the self-care of T2D (or any chronic illness). Complete audio data was recorded for all interviews except one telephone interview for which the first 10 minutes were lost due to equipment malfunction. In the 3 telephone interviews, children and partners were in the vicinity of the participant during the interview, which could have impacted the interview content. The participants were not asked to comment on the transcripts. Double coding of a subset of interviews by 2 members of the team and an ongoing discussion about the coding structure ensured that the coding scheme was robust. Multiple views of the data promote confidence in the credibility of the findings [36]. A diverse range of experiences and opposing arguments were identified and presented.

Some caution should be exercised in the transferability of the findings. Although we sought out participants from less-advantaged and more ethnically diverse neighborhoods, our sample was predominantly White and well educated. Being able to speak and understand English was used as a study entry criterion, in response to the challenges of conducting cross-language qualitative research [37] and because of the lack of availability of resources to contract an interpreting service. This may have created a barrier to study entry for some minority ethnic groups. Although not all participants were technophiles, the sample mostly included adults aged over 50 years with an interest in technology. This sample may reflect historically lower access to the internet among people from groups of lower socioeconomic status and those living in remote geographical regions [38,39].

Interpretations in the Context of Existing Literature

Most pertinent to the findings of this study is the literature on illness identities, stigma, and control of their bodies and illness [15,18,40-43]. The diagnosis of a chronic condition such as T2D has been described as an assault on the identity [14,44]. Successful adaptation is dependent on acceptance of the change to identity and associated lifestyle shifts and is mitigated by available resources [13,14,41]. In this study, participants described using digital interventions as a resource to support them to make a largely invisible disease more tangible, to help them understand their changed body, and to help them engage in self-care activities. They were using technology in a way that has been described as *digitizing the body*, providing a different way of *knowing* and *controlling* the health status of a body when physical sensations are an insufficient guide [27,32]. Digital tools increased participants' sense of control and agency, implicitly reinforcing the idea of individual responsibility for managing their illness.

The majority of participants used technology to express preferred identities and to resist giving the master status to the diabetic identity. Luttrell [45] and James [46] proposed that people tell stories that allow them to present their more desirable selves in challenging situations such as the diagnosis of a stigmatizing disease. This was reflected in this study, where participants used technology to present more desirable preferred identities, such as *data master*, rather than the stigmatized diabetic identity. Both in the interview and in their social environments, they were able to demonstrate status by sharing their superior knowledge of technology. They used this knowledge to gain power and status in their interactions with HCPs.

The experience of enacted and felt stigma following a chronic illness diagnosis has been well documented, and this study demonstrated that stigma influenced the technology participants selected [15,47]. According to the stigma theory by Goffman [47], stigma occurs with chronic illness when people behave in a way that deviates from expectations of what is *normal*. He proposed that people conceal their true identities to fit in in the world of *normals* [47]. As T2D is a relatively invisible illness, people with the condition can choose to disclose their condition, which might mean they get support but might experience stigma [17,48]. Alternatively, they can conceal the condition to avoid identifying or being identified with the stigmatized identity [17,48]. In this study, some participants were able to replace big, bulky NHS BGMs with small *discrete* alternatives, which meant they were able to choose to not be socially identified or defined by their condition [18]. Those with more money were better able to buy technology that allowed them to pass as *normal*. The desire to avoid a stigmatized diabetic identity may also explain the widespread use of fitness trackers and the limited use of diabetes-specific digital interventions.

Implications for Research and Future Intervention Development

The projection of identities through technology may be a promising route for the future development of technology to support people self-care for their T2D. In this study, there was a clear story being told by the participants with T2D about how they used technology to express positive identities and selected interventions because they supported their preferred identities. Participants in the most part did not select diabetes-specific technology and when they did, they talked about the importance of it being discreet or framed its use in terms of expression of positive identity. Some interventions have addressed identity change by providing educational modules, for example, the *Drink Less* app, which was designed to tackle excess alcohol consumption [49], and the MoveDaily intervention, which tried to link the formation of health habits to identity [31]. However, very few interventions aimed at supporting self-care of chronic conditions consider how using the interventions assists or prevents people's ability to enact positive identities, for example, by designing diabetes-specific technology (such as a BGMs) to be attractive or to appeal to other positive identities. Intervention design that highlights people's preferred identities may be more likely to be used and, therefore, have more beneficial effects.

Conclusions

This study has shown that people with T2D used technology to understand and control their body and support their preferred social identity. Digital health technology was used to support the expression of positive identities (a good patient, an expert, or in control) or to avoid a stigmatized identity (by resisting the diagnosis or hiding their condition).

Acknowledgments

The authors would like to thank the people who gave up their time to participate in this research. This project was funded by the National Institute for Health Research (NIHR) School for Primary Care Research (SPCR; project reference 2019-5007) at the University of Bristol. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Authors' Contributions

ST had primary responsibility for conducting this study as part of her NIHR SPCR PhD and drafted the manuscript. CC and PL contributed to the conception, study design, analysis, and interpretation. All authors reviewed and contributed to the revisions of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ethics favorable opinion letter.
[DOCX File , 364 KB - jmir_v22i12e21328_app1.docx]

Multimedia Appendix 2
Screening questionnaire.
[DOCX File , 2171 KB - jmir_v22i12e21328_app2.docx]

Multimedia Appendix 3
Topic guide.
[DOCX File , 1287 KB - jmir_v22i12e21328_app3.docx]

Multimedia Appendix 4
Coding tree.
[DOCX File , 24 KB - jmir_v22i12e21328_app4.docx]

Multimedia Appendix 5
Participant individual profiles.
[DOCX File , 42 KB - jmir_v22i12e21328_app5.docx]

References

1. Alonso J, Ferrer M, Gandek B, Ware JE, Aaronson NK, Mosconi P, IQOLA Project Group. Health-related quality of life associated with chronic conditions in eight countries: results from the international quality of life assessment (IQOLA) project. Qual Life Res 2004 Mar;13(2):283-298. [doi: 10.1023/b:qure.0000018472.46236.05] [Medline: 15085901]
2. The Global Burden of Disease: 2004 Update. World Health Organisation. 2004. URL: https://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/ [accessed 2018-01-10]
3. Goodwin N, Curry N, Naylor C, Ross S. Managing People With Long-Term Conditions. The Kings Fund. 2010. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_document/managing-people-long-term-conditions-gp-inquiry-research-paper-mar11.pdf [accessed 2018-01-10]
4. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. Lancet 2017 Jun 3;389(10085):2239-2251. [doi: 10.1016/S0140-6736(17)30058-2] [Medline: 28190580]
5. Type 2 Diabetes in Adults: Management. The National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/guidance/ng28 [accessed 2020-10-01]
6. Khan MA, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes - global burden of disease and forecasted trends. J Epidemiol Glob Health 2020 Mar;10(1):107-111 [FREE Full text] [doi: 10.2991/jegh.k.191028.001] [Medline: 32175717]
7. Lorig KR, Ritter P, Stewart AL, Sobel DS, Brown BW, Bandura A, et al. Chronic disease self-management program: 2-year health status and health care utilization outcomes. Med Care 2001 Nov;39(11):1217-1223. [doi: 10.1097/00005650-200111000-00008] [Medline: 11606875]
8. Supporting People and Integration With Long Term Conditions an NHS and Social Care Model to Support Local Innovation. Department of Health. 2005. URL: https://webarchive.nationalarchives.gov.uk/20130105013243/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4122574.pdf [accessed 2018-08-08]
9. Galvin R. Disturbing notions of chronic illness and individual responsibility: towards a genealogy of morals. Health (London) 2016 Jul 24;6(2):107-137. [doi: 10.1177/136345930200600201]
10. Living With Diabetes. Diabetes UK. URL: https://www.diabetes.org.uk/guide-to-diabetes [accessed 2020-10-01]
11. Ellis J, Boger E, Latter S, Kennedy A, Jones F, Foster C, et al. Conceptualisation of the 'good' self-manager: a qualitative investigation of stakeholder views on the self-management of long-term health conditions. Soc Sci Med 2017 Mar;176:25-33. [doi: 10.1016/j.socscimed.2017.01.018] [Medline: 28126586]
12. Moore L, Frost J, Britten N. Context and complexity: the meaning of self-management for older adults with heart disease. Sociol Health Illn 2015 Nov;37(8):1254-1269. [doi: 10.1111/1467-9566.12316] [Medline: 26235674]
13. Gomersall T, Madill A, Summers LK. A metasynthesis of the self-management of type 2 diabetes. Qual Health Res 2011 Jun;21(6):853-871. [doi: 10.1177/1049732311402096] [Medline: 21429946]
14. Goldman JB, Maclean HM. The significance of identity in the adjustment to diabetes among insulin users. Diabetes Educ 1998;24(6):741-748. [doi: 10.1177/014572179802400610] [Medline: 10025297]
15. Broom D, Whittaker A. Controlling diabetes, controlling diabetics: moral language in the management of diabetes type 2. Soc Sci Med 2004 Jun;58(11):2371-2382. [doi: 10.1016/j.socscimed.2003.09.002] [Medline: 15047092]
16. Adams S, Pill R, Jones A. Medication, chronic illness and identity: the perspective of people with asthma. Soc Sci Med 1997 Jul;45(2):189-201. [doi: 10.1016/s0277-9536(96)00333-4] [Medline: 9225407]
17. Hewett G. Just a Part of Me - Men's Reflections on Chronic Asthma - Occasional Papers in Sociology and Social Policy No. 1. London South Bank University. 1994. URL: https://lsbu-archive.maxarchiveservices.co.uk/index.php/just-a-part-of-me-mens-reflections-on-chronic-asthma-occasional-papers-in-sociology-and-social-policy-no-1 [accessed 2020-02-01]
18. Charmaz K. The Body, Identity, and Self: Adapting To Impairment. The Sociological Quarterly 1995 Sep;36(4):657-680. [doi: 10.1111/j.1533-8525.1995.tb00459.x]
19. Phelan JC, Link BG, Diez-Roux A, Kawachi I, Levin B. 'Fundamental causes' of social inequalities in mortality: a test of the theory. J Health Soc Behav 2004 Sep;45(3):265-285. [doi: 10.1177/002214650404500303] [Medline: 15595507]
20. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav 2010;51 Suppl:S28-S40. [doi: 10.1177/0022146510383498] [Medline: 20943581]
21. Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav 1995;Spec No:80-94. [Medline: 7560851]
22. Marmot M. Status Syndrome How Your Social Standing Directly Affects Your Health. London, UK: Bloomsbury; 2004.
23. Marmot M, Commission on Social Determinants of Health. Achieving health equity: from root causes to fair outcomes. Lancet 2007 Sep 29;370(9593):1153-1163. [doi: 10.1016/S0140-6736(07)61385-3] [Medline: 17905168]
24. Marmot M, Bell R. Fair society, healthy lives. Public Health 2012 Sep;126 Suppl 1:S4-S10. [doi: 10.1016/j.puhe.2012.05.014] [Medline: 22784581]

25. Lastrucci V, Lorini C, Caini S, Bonaccorsi G, Florence Health Literacy Research Group. Health literacy as a mediator of the relationship between socioeconomic status and health: a cross-sectional study in a population-based sample in Florence. PLoS One 2019;14(12):e0227007 [FREE Full text] [doi: 10.1371/journal.pone.0227007] [Medline: 31869381]
26. Technology and Diabetes. Diabetes UK. URL: https://www.diabetes.org.uk/guide-to-diabetes/diabetes-technology [accessed 2020-10-10]
27. Lupton D. Digitized Health Promotion: Personal Responsibility for Health in the Web 2.0 Era. Department of Sociology and Social Policy, University of Sydney. 2013. URL: https://ses.library.usyd.edu.au/bitstream/handle/2123/9190/Working%20paper%20No.%205%20-%20Digitized%20health%20promotion.pdf?sequence=1&isAllowed=y [accessed 2020-12-03]
28. Castle-Clarke S. The NHS at 70: What Will New Technology Mean for the NHS and Its Patients? King's Fund. 2018. URL: https://www.nuffieldtrust.org.uk/files/2018-06/1530028974_the-nhs-at-70-what-will-new-technology-mean-for-the-nhs-and-its-patients.pdf [accessed 2020-12-03]
29. Murray E, Burns J, See T, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005 Oct 19(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
30. Muñoz RF. Using evidence-based internet interventions to reduce health disparities worldwide. J Med Internet Res 2010 Dec 17;12(5):e60 [FREE Full text] [doi: 10.2196/jmir.1463] [Medline: 21169162]
31. Hooglugt F, Ludden GD. A mobile app adopting an identity focus to promote physical activity (MoveDaily): iterative design study. JMIR Mhealth Uhealth 2020 Jun 15;8(6):e16720 [FREE Full text] [doi: 10.2196/16720] [Medline: 32538795]
32. Armstrong N, Koteyko N, Powell J. 'Oh dear, should I really be saying that on here?': issues of identity and authority in an online diabetes community. Health (London) 2012 Jul;16(4):347-365. [doi: 10.1177/1363459311425514] [Medline: 22067915]
33. Strauss A, Corbin JM. Grounded Theory in Practice. California, United States: SAGE Publications Inc; 1997.
34. Thomas DR. A General Inductive Approach For Qualitative Data Analysis. Frankumstein. 2003 Aug. URL: https://frankumstein.com/PDF/Psychology/Inductive%20Content%20Analysis.pdf [accessed 2020-10-01]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Sandelowski M. Sample size in qualitative research. Res Nurs Health 1995 Apr;18(2):179-183. [doi: 10.1002/nur.4770180211] [Medline: 7899572]
37. Squires A. Methodological challenges in cross-language qualitative research: a research review. Int J Nurs Stud 2009 Feb;46(2):277-287 [FREE Full text] [doi: 10.1016/j.ijnurstu.2008.08.006] [Medline: 18789799]
38. Hardiker NR, Grant MJ. Factors that influence public engagement with eHealth: a literature review. Int J Med Inform 2011;80(1):1-12. [doi: 10.1016/j.ijmedinf.2010.10.017] [Medline: 21112244]
39. Gibbons MC. A historical overview of health disparities and the potential of eHealth solutions. J Med Internet Res 2005 Oct 4;7(5):e50 [FREE Full text] [doi: 10.2196/jmir.7.5.e50] [Medline: 16403714]
40. Whitehead M, Pennington A, Orton L, Nayak S, Petticrew M, Sowden A, et al. How could differences in 'control over destiny' lead to socio-economic inequalities in health? a synthesis of theories and pathways in the living environment. Health Place 2016 May;39:51-61 [FREE Full text] [doi: 10.1016/j.healthplace.2016.02.002] [Medline: 26986982]
41. Riessman CK. Performing identities in illness narrative: masculinity and multiple sclerosis. Qual Res 2016 Aug 15;3(1):5-33. [doi: 10.1177/146879410300300101]
42. Bury M. The sociology of chronic illness: a review of research and prospects. Sociol Health Illn 1991 Dec;13(4):451-468. [doi: 10.1111/j.1467-9566.1991.tb00522.x]
43. Bury M. Illness narratives: fact or fiction? Sociol Health Illn 2001 Dec 23;23(3):263-285. [doi: 10.1111/1467-9566.00252]
44. Bury M. Chronic illness as biographical disruption. Sociol Health Illn 1982 Jul;4(2):167-182 [FREE Full text] [doi: 10.1111/1467-9566.ep11339939] [Medline: 10260456]
45. Luttrell W. Pregnant Bodies, Fertile Minds: Gender, Race, and the Schooling of Pregnant Teens. Abingdon, United Kingdom: Routledge; Jun 25, 2003.
46. James D. Dona Maria's Story: Life History, Memory, and Political Identity. Durham, North Carolina: Duke University Press; 2000.
47. Goffman E. Stigma: Notes on the Management of Spoiled Identity. New York City, USA: Simon and Schuster; 2009.
48. Joachim G, Acorn S. Stigma of visible and invisible chronic conditions. J Adv Nurs 2000 Jul;32(1):243-248. [doi: 10.1046/j.1365-2648.2000.01466.x] [Medline: 10886457]
49. Garnett C, Crane D, Michie S, West R, Brown J. Evaluating the effectiveness of a smartphone app to reduce excessive alcohol consumption: protocol for a factorial randomised control trial. BMC Public Health 2016 Jul 8;16:536 [FREE Full text] [doi: 10.1186/s12889-016-3140-8] [Medline: 27392430]

Abbreviations

BGM: blood glucose monitor
HCP: health care practitioner

NHS: National Health Service
NIHR: National Institute for Health Research
T2D: type 2 diabetes
SPCR: School for Primary Care Research

Edited by A Moorhead; submitted 11.06.20; peer-reviewed by S Hughes, M Harris; comments to author 09.07.20; revised version received 28.10.20; accepted 17.11.20; published 21.12.20.

Please cite as:
Turnbull S, Lucas PJ, Hay AD, Cabral C
Digital Health Interventions for People With Type 2 Diabetes to Develop Self-Care Expertise, Adapt to Identity Changes, and Influence Other's Perception: Qualitative Study
J Med Internet Res 2020;22(12):e21328
URL: http://www.jmir.org/2020/12/e21328/
doi:10.2196/21328
PMID:33346733

Original Paper

Feasibility and Acceptability of a Remotely Delivered, Web-Based Behavioral Intervention for Men With Prostate Cancer: Four-Arm Randomized Controlled Pilot Trial

June M Chan[1*], ScD; Erin L Van Blarigan[1*], ScD; Crystal S Langlais[1], MPH; Shoujun Zhao[1], PhD; Justin W Ramsdill[2], MS; Kimi Daniel[2], MS, ATC; Greta Macaire[1], MA, RD, CSO; Elizabeth Wang[3], MPH; Kellie Paich[4], MPH; Elizabeth R Kessler[5], MD; Tomasz M Beer[2], MD; Karen S Lyons[6], PhD; Jeanette M Broering[1], RN, PhD, MPH; Peter R Carroll[1], MD, MPH; Stacey A Kenfield[1*], ScD; Kerri M Winters-Stone[2*], PhD

[1]University of California, San Francisco, San Francisco, CA, United States
[2]Oregon Health and Science University, Portland, OR, United States
[3]Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, United States
[4]Movember Foundation, Culver City, CA, United States
[5]University of Colorado, Aurora, CO, United States
[6]Boston College, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Erin L Van Blarigan, ScD
University of California, San Francisco
UCSF Box 0560
550 16th St. 2nd Floor
San Francisco, CA, 94158
United States
Phone: 1 415 476 1111 ext 13608
Email: erin.vanblarigan@ucsf.edu

Abstract

Background: Diet and exercise may be associated with quality of life and survival in men with prostate cancer.

Objective: This study aimed to determine the feasibility and acceptability of a remotely delivered web-based behavioral intervention among men with prostate cancer.

Methods: We conducted a multi-site 4-arm pilot randomized controlled trial of a 3-month intervention (TrueNTH Community of Wellness). Eligibility included self-reported prostate cancer diagnosis, having a personal device that connected to the internet, age ≥18 years, and ability to read English and receive text messages and emails. Men receiving chemotherapy or radiation, or those who reported contraindications to exercise, could participate with physician clearance. Participants were randomized (1:1:1:1) to additive intervention levels: website; website and personalized diet and exercise prescription; website, personalized prescription, Fitbit, and text messages; and website, personalized prescription, Fitbit, text messages, and 2 30-minute phone calls—one with an exercise trainer and one with a registered dietician. Primary outcomes were feasibility (accrual and attrition) and acceptability (survey data and website use). We described self-reported diet and exercise behavior at the time of enrollment, 3 months, and 6 months as secondary outcomes.

Results: In total, 202 men consented and were randomized between August 2017 and September 2018 (level 1: 49, level 2: 51, level 3: 50, level 4: 52). A total of 160 men completed the onboarding process and were exposed to their randomly assigned intervention (38, 38, 42, and 42 in levels 1, 2, 3, and 4, respectively). The follow-up rate was 82.7% (167/202) at 3 months and 77.2% (156/202) at 6 months. Participants had a median age of 70 years and were primarily White and college educated. Website visit frequency over the 3-month intervention period increased across levels (median: 2, 9, 11, and 16 visits for levels 1, 2, 3, and 4, respectively). Most were satisfied or very satisfied with the intervention (20/39, 51%; 27/42, 64%; 23/44, 52%; and 27/42, 64% for levels 1, 2, 3, and 4, respectively). The percentage of men who reported being very satisfied was highest among level 4 participants (10/42, 24% vs 4/39, 10%; 5/42, 12%; and 5/44, 11% for levels 1, 2, and 3, respectively). Dissatisfaction was highest in level 1 (5/39, 13% vs 1/42, 2%; 3/44, 7%; and 2/42, 5% for levels 2, 3, and 4, respectively). We observed small improvements in diet and physical activity at 3 months among men in level 4 versus those in level 1.

Conclusions: A web-based, remotely delivered, tailored behavioral intervention for men with prostate cancer is feasible. Future studies are warranted to increase the effect of the intervention on patient behavior while maintaining sustainability and scalability as well as to design and implement interventions for more diverse populations.

Trial Registration: ClinicalTrials.gov NCT03406013; http://clinicaltrials.gov/ct2/show/NCT03406013

(J Med Internet Res 2020;22(12):e19238) doi:10.2196/19238

KEYWORDS

diet; physical activity; exercise; lifestyle; cancer; survivorship; text messages; internet

Introduction

Background

Prostate cancer affects more than 3.6 million men in the United States, making it the most prevalent cancer in American men [1]. Prostate cancer is characterized by low age-adjusted death rates and a relatively long median survival time of 16 years, although this varies greatly by stage at diagnosis [2]. Over this period, men with prostate cancer may experience significant disease- or treatment-related decline in quality of life [3-6], including incontinence, erectile dysfunction, fatigue, poor metabolic functioning, reduced bone and muscle integrity, hot flushes, sexual dysfunction, and low mood [7].

Prior Work

A growing number of studies have suggested the benefits of a healthy diet and regular exercise for men with prostate cancer, including lower risk of treatment-associated side effects [8-11], cancer progression [12], and cancer-specific mortality [13-16]. Specific dietary factors that have been associated with improved clinical outcomes in men with prostate cancer include higher intake of cruciferous vegetables, vegetable fat, fish, and cooked tomatoes and lower intake of processed meat, whole milk, and poultry with skin [17-25]. Physical activity recommendations of ≥150 minutes of moderate intensity or ≥75 minutes of high-intensity aerobic exercise per week have also been associated with a lower risk of mortality in men with prostate cancer [26].

Translation of this growing evidence of possible benefits of a healthy lifestyle for prostate cancer survivors into clinical practice and survivorship programs has been slow. Although physical activity and nutrition guidelines exist for cancer survivors [27], lifestyle counseling and exercise programs are not standard care practices for individuals with cancer in the United States. Moreover, most prostate cancer survivors do not follow the recommendations.

Previous studies on lifestyle interventions suggest that center-based interventions are effective but require infrastructure that may only be available in urban academic centers [28]. In contrast, it is challenging to make home-based interventions comprehensive or tailored to the individual needs and interests of the participants [29]. The diverse attitudes and motivations to change behavior among men with prostate cancer further complicate matters. Previous studies suggest that although men with prostate cancer appreciate the importance of exercise, most do not feel that the information provided by their doctor is specific enough to be actionable [30]; many also report low motivation for physical activity [30].

Objectives

In this context, web-based interventions are emerging as promising, scalable modalities for behavior change [31-36]. Prior literature suggests that tailoring of an intervention to individual characteristics and goals and combining technology with personal guidance (ie, blended intervention) may lead to improved outcomes [37-39]. However, questions regarding the feasibility and acceptability of remotely delivered web-based interventions remain, particularly in older adult populations. Thus, we designed the TrueNTH Community of Wellness study, a 4-arm, multi-site, pilot randomized controlled trial to evaluate the feasibility and acceptability of a 3-month web-based intervention for men with prostate cancer with progressive levels of behavioral support. In this study, we report our primary results on the feasibility and acceptability of the intervention. Second, we describe self-reported levels of physical activity and diet at the time of enrollment and 3 and 6 months after the enrollment and explore changes in lifestyle behaviors over the study period.

Methods

Study Design, Population, and Recruitment

We conducted a 4-arm pilot randomized controlled trial of a 3-month intervention among men with prostate cancer (ClinicalTrials.gov NCT03406013) to compare 3 levels of increasing behavioral support (levels 2-4) with general educational information on a website (level 1). The multicenter trial was conducted at Oregon Health Sciences University (OHSU), University of California San Francisco (UCSF), and University of Colorado Denver (UCD).

The trial protocol has been reported [40]. Briefly, men were recruited through hospital cancer registry databases, the Cancer of the Prostate Strategic Urologic Research Endeavor registry of men with prostate cancer, and in clinics. Men were eligible to participate if they self-reported a prostate cancer diagnosis, had a personal device with internet and text messaging capabilities and a personal email address, were aged ≥18 years, and were able to read English. Men currently receiving chemotherapy or radiation therapy or those who had potential contraindications to exercise identified on the basis of the

American College of Sports Medicine exercise preparticipation screening criteria could participate with physician clearance.

OHSU was the primary coordinating center for this trial. All participants provided written consent, and all study-related activities were performed in accordance with and under the supervision of the institutional review board of each study site.

Randomization and Blinding

Consenting men were block randomized (1:1:1:1) by site (UCSF, OHSU, or UCD) to increasing levels of web-based behavioral support (level 1-4). The randomization scheme was computer generated with block size 4 by SZ and stored at UCSF. When a participant had completed consent and enrollment procedures, research staff at OHSU requested the next assignment from UCSF. Participants were told that they would be randomly assigned to different tools and resources but were unaware of which resources they received relative to other participants.

Intervention Description

Details on the intervention, including its theoretical basis, have been previously reported [40]. Briefly, level 1 (reference group) received general educational information about exercise and diet, a resource directory, and study-specific guidelines delivered through the website (information on the website was not changed throughout the study period for level 1). Level 2 received the information provided to level 1 along with a personalized diet and exercise prescriptions delivered through the website, including videos of recommended exercises and a weekly short survey about their progress toward the diet and exercise recommendations. Level 3 received information and resources provided to level 2 along with a Fitbit Alta (Fitbit Inc) with physical activity reports (Fitbit data integrated into the website), supportive text messages (50 texts over 90 days: average 4 per week, no response required, equally split and alternating between diet and exercise topics), and weekly web-based short surveys for participants to track their progress toward the diet and exercise recommendations. Level 4 received information and resources provided to level 3 as well as 2 optional 30-minute calls: one with an exercise trainer (KD) and one with a registered dietician (GM). Men accessed the study website with a username and password and had to complete an onboarding process to gain access to the intervention website. The study website home page included a dashboard that summarized self-reported diet and physical activity behavior using visuals and contained links to other pages (eg, Report Progress, See Progress, Connect Fitbit, etc), depending on the assigned level. All participants from all levels received an instruction sheet (PDF) orienting them to the website at enrollment and a weekly email reminder to encourage them to use the website.

The intervention recommendations focused on diet and physical activity. The dietary recommendations provided to all levels were to consume one serving each of healthy vegetable fats and cruciferous vegetables per day; 2 servings each of cooked tomatoes and fish per week; and no whole milk, processed meat, or poultry with skin. The individual dietary prescriptions provided to levels 2 to 4 were focused on helping the participants achieve these recommendations, considering what the participants self-reported at baseline. The physical activity recommendations, consistent with national guidelines, were to engage in 150 minutes of aerobic exercise per week, 60 minutes of strength training per week accumulated in ≥2 sessions, and 2 sessions of stretching per week. The individual exercise prescriptions provided to levels 2 to 4 were based on the men's self-reported current physical activity levels, health goals for exercise, health status, resources for exercise, and time available for exercise. In addition, the exercise prescriptions were tailored to participants' self-reported current prostate cancer status. Men with bone metastases or active cancer treatment other than androgen deprivation therapy were prescribed low-intensity programs.

Web-Based Survey Assessments

Participants were asked to complete surveys on the internet at baseline, 3 months (immediately following the intervention), and 6 months using the study website and Research Electronic Data Capture [41]. Surveys included sociodemographics and prostate cancer diagnosis and treatment (baseline only), the Community Health Activities Model Program for Seniors (CHAMPS) [42], and a validated food frequency questionnaire (FFQ) [43]. Each month, the men also completed a web-based survey about adverse events (AEs) and whether any reported AE was related to a pre-existing condition present before enrollment. After 3 months, men were asked to complete a web-based acceptability survey specific to the level of intervention received.

Primary Outcomes: Feasibility and Acceptability

The primary study outcomes were feasibility and acceptability of the intervention. Feasibility was assessed based on the accrual time and retention. Accrual was defined as the time between enrollment of the first and last participant; our goal was to accrue 200 participants in 12 months. Retention was defined as the proportion of participants who completed at least one of the surveys at each follow-up time point (3 and 6 months). A priori, we specified that we would consider the intervention to be feasible if we retained at least 80% (160/200) of participants at 3 months and 64% (128/200) of participants at 6 months [40].

Acceptability was measured via an investigator-developed web-based survey administered at 3 months and website use. The level-specific survey asked men to assess their overall satisfaction with the program as well as each of the resources received (website, resource directory, exercise prescription, diet prescription, weekly progress report, Fitbit, text messages, diet coach, and exercise coach) as *very satisfied*, *satisfied*, *neutral*, *dissatisfied*, *very dissatisfied*, or *did not use*. To assess overall satisfaction, participants were asked "How satisfied were you with your experience with the Community of Wellness intervention and portal?" When assessing resources, participants were only asked about resources they received based on their assigned level of intervention. There was also an open text box for additional comments and feedback. The frequency of visits to the study portal was directly assessed using log-ins and web analytic data. A priori, we expected all men to log onto the website at least once and that the frequency of log-ins would increase across levels from level 1 to 4 [40]. There was a delay in activating the analytics function between August and December 2017; therefore, there were no data for the first 7 enrollees and partial data for those enrolled between October and December 2017. Thus, summaries of web analytics data reflect an underestimate of use.

Secondary Outcomes: Change and Maintenance of Diet and Physical Activity

The intervention was designed to support the adoption or maintenance of diet and physical activity habits previously reported to be associated with reducing the risk of prostate cancer recurrence or mortality. Thus, we estimated the effect of the intervention levels on changes in self-reported diet and physical activity as secondary outcomes.

Diet Assessment

We used a validated FFQ to assess the usual diet at the time of enrollment and 3 and 6 months after the enrollment [43]. The FFQ asked men to report their usual intake of approximately 140 foods and beverages over the past 3 months, with up to nine frequency options ranging from never or <1 per month to ≥6 per day. *Cruciferous vegetables* included a half cup of broccoli, cauliflower, cabbage, brussels sprouts, kale, mustard greens, or chard. *Vegetable fat* included avocado (half cup), oil dressing (1-2 tbs), peanut butter (1 tbs), peanuts (1 oz), walnuts (1 oz), and other nuts (1 oz). *Cooked tomatoes* included intake of tomato sauce (half cup), salsa (quarter cup), and pizza (2 slices). *Fish* included tuna fish (3-4 oz), fish sticks (1 serving), dark meat fish (eg, mackerel, salmon, sardines, bluefish, swordfish; 3-5 oz), and other fish (3-5 oz). *Processed meat* included intake of beef or chicken hotdogs (1 hotdog), bacon (2 slices), processed meat sandwiches (eg, salami, bologna), and other processed meat (eg, sausage, kielbasa; 2 oz or 2 small links). Consumption of *whole milk* was assessed by asking men how frequently they consumed 1 glass (8 oz) of whole milk. Consumption of *poultry with skin* was assessed by asking men how frequently they consumed 3 to 4 oz of chicken or turkey with skin. FFQs with >70 items blank were considered incomplete (31 participants at 3 months and 36 participants at 6 months); these data were not included in secondary analyses examining dietary change.

Physical Activity Assessment

We used the CHAMPS survey to quantify 3 types of physical activities: aerobic, strength training, and stretching. *Aerobic activity* included minutes per week (min/wk) of moderate-to-vigorous aerobic activities, including tennis, skating (ice, roller, or inline), jogging or running, walking or hiking uphill, walking fast for exercise, riding bikes, aerobic machines (eg, rowing, step), swimming, water exercises, aerobics or dancing, and sports (eg, basketball, soccer, racquetball). We calculated the number of sessions and total time per week spent doing *strengthening activities* and the number of weekly *stretching* sessions, as reported in the CHAMPS survey [42]. If a participant responded to at least one item on the CHAMPS survey, items that were skipped were assumed to be zero. If a participant did not respond to any question, his physical activity was missing for that time point (3 men at baseline, 49 men at 3 months, and 57 men at 6 months).

Total Lifestyle Behavior Score

To combine diet and physical activity into one measure, we developed a total lifestyle behavior score [40]. Multimedia Appendix 1 lists the items and points assigned to calculate the score. Men were given points for each component: 0 (not meeting recommendation), 1 (almost meeting recommendation), or 2 (meeting recommendation). Scores were summed across components to develop a diet score ranging from 0 to 14, with 14 assigned to men fully meeting dietary recommendations; a physical activity score ranging from 0 to 6, with 6 assigned to men fully meeting the physical activity recommendations; and an overall lifestyle score ranging from 0 to 20, with a score of 20 assigned to men fully meeting all lifestyle recommendations. One recommendation (ie, only taking supplements that have been reviewed with a health professional) was unable to be included in the score [40], as the question was inadvertently omitted from the 3-month survey following a switch in the technology platform after the initial 20 participants.

Statistical Analysis

Descriptive statistics, including proportions for categorical variables and median (IQR) for continuous variables, were used to summarize the characteristics of the participants, overall and by level of intervention received. We also used descriptive statistics to describe responses to the acceptability survey. Two participants originally randomized to level 1 and level 2 were incorrectly provided access to the level 3 intervention. In our primary acceptability and feasibility analyses, we analyzed these individuals based on intervention received (level 3), as acceptability surveys and web analytics were specific to level received. To compute participation proportions and secondary analyses of behavior change and AEs, the men were analyzed using the original randomization level (level 1 and level 2).

To explore diet and physical activity behaviors, we calculated the median (IQR) dietary score, physical activity score, and total lifestyle behavior score at each time point. Summary statistics for each component of diet and physical activity scores per arm of the study at each time point were also reported. We used *t* tests to calculate the mean change and 95% CI between baseline and 3 months and baseline and 6 months for each of the scores, within and between arms. As these were secondary analyses, we used a complete case approach and described levels of diet and physical activity at each time point among participants with available data.

In a posthoc sensitivity analysis, we also examined aerobic activity at each time point by level and aerobic physical activity at enrollment. We hypothesized that men with low levels of physical activity at enrollment would increase their activity at 3 months, whereas there would be no change among men who were already performing the recommended 150 minutes per week of physical activity.

Analyses were performed using Stata version 16.0 (StataCorp LP). As specified in our protocol, we tested for differences in change in the 3 behavior scores across arms using the standard α level of .05 to assess statistical significance. For the rest of our analyses, we adhered to the Consolidated Standards of

Reporting Trials recommendation for pilot and feasibility studies and report descriptive statistics only [44].

Results

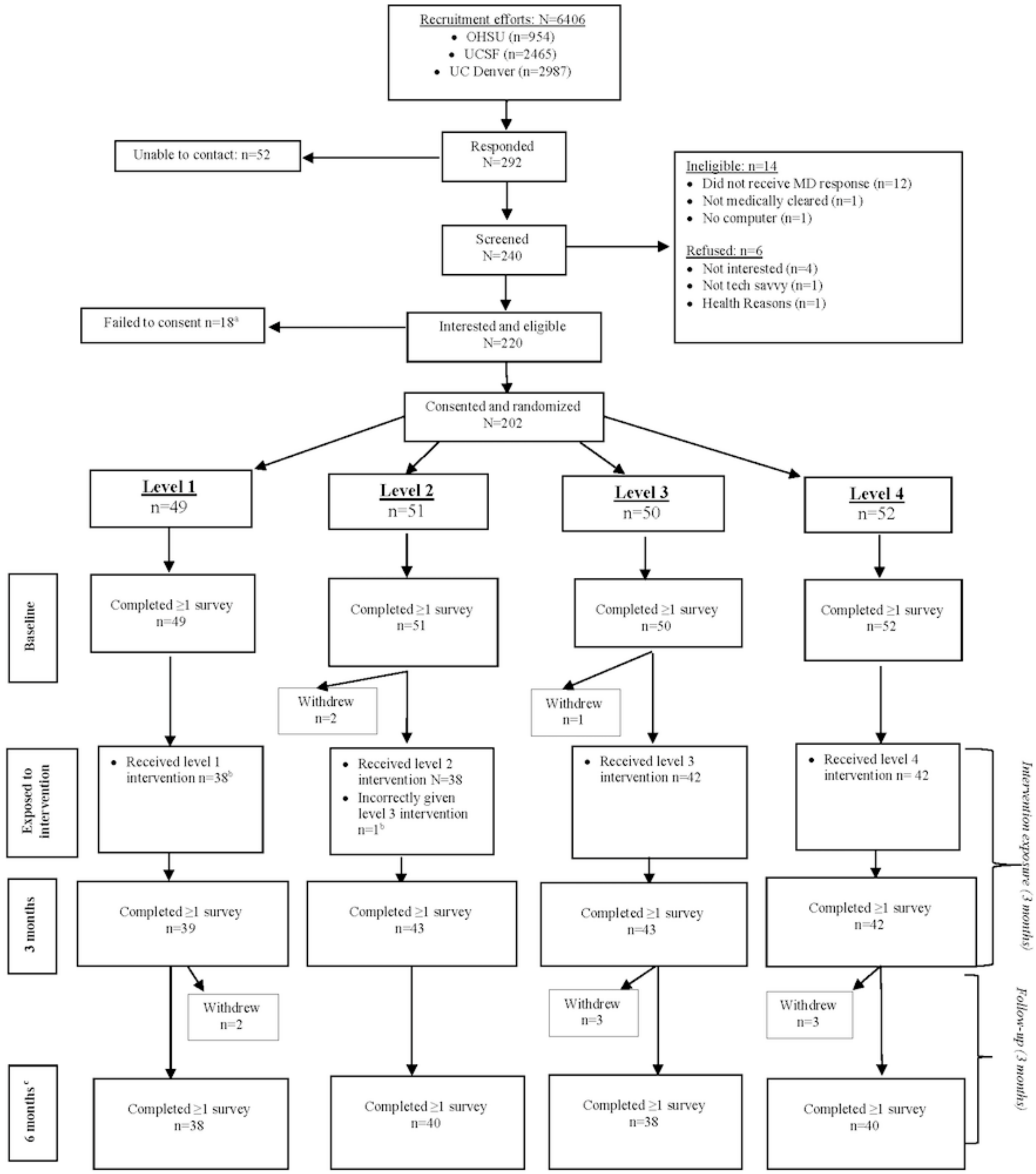
Recruitment and Retention

A total of 6406 men received a letter, between July 2017 and September 2018, detailing the study and providing contact information for those who wanted to learn more (Figure 1). Of these, 292 men expressed interest, 240 were screened for eligibility, and 220 were interested and deemed eligible. The main reason for ineligibility was not receiving medical clearance (13 men). Furthermore, 1 man did not own a computer and 6 declined to participate after screening. Of the 220 interested and eligible men, 202 were randomized and provided access to the web-based consenting process and surveys. These 202 men comprise the initial population for analysis.

The 202 men with prostate cancer were randomized (1:1:1:1) to increasing levels of web-based behavioral support; 49 were assigned to level 1, 51 were assigned to level 2, 50 were assigned to level 3, and 52 were assigned to level 4 (Figure 1). Of the 202 randomized men, 161 completed the onboarding process and were exposed to the intervention. Of these men, 160 were exposed to their randomly assigned intervention (38 in level 1, 38 in level 2, 42 in level 3, and 42 in level 4). As noted above, 2 men were incorrectly given access to level 3, when originally assigned to levels 1 and 2; only one of these men (originally assigned to level 2) initiated the intervention by accessing the study website. Throughout the 6-month follow-up, 11 men withdrew after randomization; the primary reason for withdrawal was the time commitment. Withdrawal was similar across groups; 2 men withdrew in level 1, 2 men withdrew in level 2, 4 men withdrew in level 3, and 3 men withdrew in level 4.

Overall, 82.7% (167/202) and 77.2% (156/202) of men completed surveys after 3 and 6 months, respectively. By level, the 3- and 6-month follow-up proportions were 80% (39/49) and 78% (38/ 49) for level 1, 84% (43/ 51) and 78% (40/ 51) for level 2, 86% (43/ 50) and 76% (38/50) for level 3, and 81% (42/52) and 77% (40/52) for level 4 (Figure 1).

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram showing the flow of participants from screening through end of study. MD: medical doctor; OHSU: Oregon Health and Sciences University; UC Denver: University of Colorado Denver; UCSF: University of California San Francisco.



[a] Three of the 18 excluded individuals completed surveys and were randomized and provided access to the intervention before noting that they did not complete online consent; they were subsequently withdrawn from the study population. [b] Men completed an online onboarding process to receive access to the intervention. Two men (one randomized to Level 1 and one randomized to Level 2) completed onboarding but were incorrectly given access to the Level 3 intervention; of these, the level 1 man never accessed the intervention and the level 2 man accessed the level 3 intervention. [c] Completing the 6m survey was not dependent on completing the 3m survey.

Characteristics of the Study Population

Of the 202 randomized participants, the men were predominantly White (187/202, 92.6%), well-educated (167/202, 82.7%, reported a 4-year college degree or more), and married or had a partner (185/202, 91.6%; Table 1). The median (IQR) age at enrollment was 70 years (65-75 years) and the median BMI was 27 kg/m^2 (25-29 kg/m^2). Participants self-reported a wide spectrum of prostate cancer stages and grades. The median (IQR) prostate-specific antigen at diagnosis was 6 ng/ml (5-10 ng/ml), 13.9% (28/202) reported T3 disease, and 5.9% (12/202) reported T4 disease; 39.6% (80/202) had intermediate grade (Gleason sum 7) and 22.8% (46/202) had high grade (Gleason sum 8-10) cancer. Most participants reported that they had been treated for localized disease at the time of enrollment or were on active surveillance. The characteristics of the men assigned to each of the 4 levels were similar.

Table 1. Baseline self-reported sociodemographic and clinical characteristics of 202 men with prostate cancer participating in a technology-supported physical activity and dietary intervention, overall and by level randomized.

Characteristics[a]	Level 1	Level 2	Level 3	Level 4	Total
Number of men, n (%)	49 (24.3)	51 (25.2)	50 (24.8)	52 (25.7)	202 (100.0)
Age (years), median (IQR)	70 (64-76)	70 (64-75)	70 (64-75)	70 (65-74)	70 (65-75)
BMI (kg/m^2), median (IQR)	25 (23-28)	28 (26-30)	26 (24-29)	27 (25-29)	27 (25-29)
Race, n (%)					
White	44 (90)	45 (88)	48 (96)	50 (96)	187 (92.6)
Black	2 (4)	3 (6)	0 (0)	0 (0)	5 (2.5)
Other	0 (0)	1 (2)	0 (0)	0 (0)	1 (0.5)
Asian	1 (2)	0 (0)	0 (0)	1 (2)	2 (1.0)
More than one race	2 (4)	2 (4)	0 (0)	1 (2)	5 (2.5)
Decline to answer	0 (0)	0 (0)	2 (4)	0 (0)	2 (1.0)
Education, n (%)					
Grade school	0 (0)	1 (2)	0 (0)	0 (0)	1 (0.5)
High school	4 (8)	6 (12)	1 (2)	3 (6)	14 (6.9)
2-year college	9 (18)	4 (8)	2 (4)	5 (10)	20 (9.9)
4-year college	10 (20)	14 (27)	20 (40)	14 (27)	58 (28.7)
Graduate or professional degree	26 (53)	26 (51)	27 (54)	30 (58)	109 (54.0)
Employment, n (%)					
Decline to answer	0 (0)	0 (0)	0 (0)	1 (2)	1 (0.5)
Full time	12 (24)	11 (22)	17 (34)	10 (19)	50 (24.8)
Part time	6 (12)	2 (4)	5 (10)	8 (15)	21 (10.4)
Retired	31 (63)	37 (73)	27 (54)	31 (60)	126 (62.4)
Disabled	0 (0)	0 (0)	0 (0)	1 (2)	1 (0.5)
Unemployed	0 (0)	1 (2)	1 (2)	1 (2)	3 (1.5)
Marital status, n (%)					
Married or partnered	46 (94)	46 (90)	46 (92)	47 (90)	185 (91.6)
Divorced or separated	1 (2)	1 (2)	1 (2)	3 (6)	6 (3.0)
Widowed	1 (2)	2 (4)	2 (4)	0 (0)	5 (2.5)
Never married	0 (0)	2 (4)	1 (2)	2 (4)	5 (2.5)
Decline to answer	1 (2)	0 (0)	0 (0)	0 (0)	1 (0.5)
PSA[b] at Dx[c] (ng/ml), median (IQR)	5 (4-10)	7 (5-11)	7 (5-10)	6 (5-12)	6 (5-10)
T-stage[d], n (%)					
T1	15 (31)	15 (30)	13 (26)	19 (37)	62 (30.7)
T2	25 (51)	19 (38)	29 (58)	14 (27)	87 (43.1)
T3	5 (10)	12 (24)	5 (10)	6 (12)	28 (13.9)
T4	1 (2)	2 (4)	3 (6)	6 (12)	12 (5.9)
Unknown	3 (6)	2 (4)	0 (0)	7 (13)	12 (5.9)
Gleason sum at Dx, n (%)					
5, low grade	0 (0)	0 (0)	1 (2)	3 (6)	4 (2.0)
6, low grade	6 (12)	7 (14)	13 (26)	8 (15)	34 (16.8)
3+4, intermediate grade	16 (33)	5 (10)	10 (20)	10 (19)	41 (20.3)

Characteristics[a]	Level 1	Level 2	Level 3	Level 4	Total
4+3, intermediate grade	10 (20)	12 (24)	11 (22)	6 (12)	39 (19.3)
8-10, high grade	8 (16)	14 (28)	12 (24)	12 (23)	46 (22.8)
Unknown or Do not know	9 (18)	12 (24)	3 (6)	13 (25)	37 (18.3)
Disease status at enrollment[e], n (%)					
On active surveillance, PSA low or not-rising	3 (6)	7 (14)	3 (6)	6 (12)	19 (9.4)
On active surveillance, PSA elevated or rising	1 (2)	2 (4)	5 (10)	4 (8)	12 (5.9)
Completed treatment for localized disease, PSA low or undetectable	31 (63)	23 (45)	26 (52)	27 (52)	107 (53.0)
Completed treatment for localized disease, PSA elevated	3 (6)	1 (2)	4 (8)	2 (4)	10 (5.0)
Cancer spread locally	1 (2)	2 (4)	1 (2)	3 (6)	7 (3.5)
Metastatic disease	2 (4)	5 (10)	4 (8)	3 (6)	14 (6.9)
Other or Do not know	8 (16)	11 (22)	7 (14)	7 (13)	33 (16.3)
Time since Dx (years)[f], median (IQR)	9 (4-14)	4 (2-7)	3 (2-10)	4 (2-9)	4 (2-10)
Time since first treatment (years), median (IQR)	2 (1-9)	4 (2-10)	3 (1-11)	3 (1-6)	3 (1-9)
Enrollment site, n (%)					
Oregon (OHSU[g])	21 (43)	23 (45)	15 (30)	18 (35)	77 (38.1)
California (UCSF[h])	16 (33)	14 (27)	19 (38)	19 (37)	68 (33.7)
Colorado (UCD[i])	12 (24)	14 (27)	16 (32)	15 (29)	57 (28.2)

[a] Percentages may not sum to 100% due to rounding.

[b] PSA: prostate-specific antigen.

[c] Dx: diagnosis.

[d] T-Stage: Tumor (T) component of the TNM staging system developed by the American Joint Committee on Cancer.

[e] Self-reported disease status.

[f] Diagnosis date was known for 127 men (31, 30, 32, 34 in levels 1-4), and the first treatment date was known for 135 men (26, 34, 39, 36 in levels 1-4).

[g] OHSU: Oregon Health Sciences University.

[h] UCSF: University of California San Francisco.

[i] UCD: University of Colorado Denver.

Acceptability of Intervention

As hypothesized, engagement and satisfaction with the intervention increased across levels (Table 2). On the basis of web analytics data on 154 of 161 men exposed to the intervention, the median (IQR) number of days that participants went to the study website over the 3-month intervention was 2 (IQR 1-3) in level 1, 9 (IQR 5-13) in level 2, 11 (IQR 8-16) in level 3, and 16 (IQR 10-19) in level 4. The most commonly visited pages for men in levels 2 to 4 were the main home page/dashboard and the follow-up survey page where participants could self-report progress toward their diet and exercise goals. Although all other pages were visited at least once, the median number of days that participants visited the remaining pages was low (1-5 visits each).

Table 2. Acceptability of a 3-month technology-supported behavioral intervention for men with prostate cancer by level of intervention received.

Characteristics[a]	Level 1	Level 2	Level 3	Level 4
Number of men[b]	39	42	44	42
Frequency of portal visits (days), median (IQR)[b]	2 (1-3)	9 (5-13)	11 (8-16)	16 (10-19)
Overall satisfaction with Community of Wellness intervention and portal, n (%)				
Very satisfied	4 (10)	5 (12)	5 (11)	10 (24)
Satisfied	16 (41)	22 (52)	18 (41)	17 (40)
Neutral	10 (26)	11 (26)	15 (34)	11 (26)
Dissatisfied	4 (10)	1 (2)	2 (5)	1 (2)
Very dissatisfied	1 (3)	0 (0)	1 (2)	1 (2)
Did not report	4 (10)	3 (7)	3 (7)	2 (5)
Resource directory, n (%)				
Excellent	1 (3)	2 (5)	0 (0)	4 (10)
Very good	5 (13)	9 (21)	4 (9)	6 (14)
Good	12 (31)	9 (21)	14 (32)	8 (19)
Average	3 (8)	3 (7)	5 (11)	7 (17)
Poor	0 (0)	2 (5)	0 (0)	3 (7)
Did not use	13 (33)	12 (29)	18 (41)	10 (24)
Did not report	5 (13)	5 (12)	3 (7)	4 (10)
Exercise prescription, n (%)				
Excellent	—[c]	1 (2)	3 (7)	10 (24)
Very good	—	11 (26)	6 (14)	10 (24)
Good	—	13 (31)	19 (43)	10 (24)
Average	—	5 (12)	6 (14)	5 (12)
Poor	—	0 (0)	0 (0)	1 (2)
Did not use	—	4 (10)	6 (14)	4 (10)
Did not report	—	8 (19)	4 (9)	2 (5)
Diet prescription, n (%)				
Excellent	—	2 (5)	6 (14)	12 (29)
Very good	—	11 (26)	11 (25)	12 (29)
Good	—	12 (29)	12 (27)	4 (10)
Average	—	5 (12)	3 (7)	8 (19)
Poor	—	0 (0)	2 (5)	0 (0)
Did not use	—	4 (10)	6 (14)	3 (7)
Did not report	—	8 (19)	4 (9)	3 (7)
Weekly progress report, n (%)				
Excellent	—	2 (5)	3 (7)	11 (26)
Very good	—	15 (36)	12 (27)	13 (31)
Good	—	6 (14)	14 (32)	6 (14)
Average	—	4 (10)	7 (16)	6 (14)
Poor	—	1 (2)	1 (2)	1 (2)
Did not use	—	6 (14)	3 (7)	2 (5)
Did not report	—	8 (19)	4 (9)	3 (7)

Characteristics[a]	Level 1	Level 2	Level 3	Level 4
Fitbit, n (%)				
Excellent	—	—	9 (20)	6 (14)
Very good	—	—	5 (11)	8 (19)
Good	—	—	7 (16)	7 (17)
Average	—	—	3 (7)	1 (2)
Poor	—	—	8 (18)	4 (10)
Did not use	—	—	7 (16)	12 (29)
Did not report	—	—	5 (11)	4 (10)
Text messages, n (%)				
Excellent	—	—	0 (0)	4 (10)
Very good	—	—	4 (9)	8 (19)
Good	—	—	14 (32)	8 (19)
Average	—	—	10 (23)	2 (5)
Poor	—	—	1 (2)	5 (12)
Did not use	—	—	10 (23)	11 (26)
Did not report	—	—	5 (11)	4 (10)
Diet coach, n (%)				
Excellent	—	—	—	9 (21)
Very good	—	—	—	9 (21)
Good	—	—	—	6 (14)
Average	—	—	—	6 (14)
Poor	—	—	—	1 (2)
Did not use	—	—	—	7 (17)
Did not report	—	—	—	4 (10)
Exercise coach, n (%)				
Excellent	—	—	—	8 (19)
Very good	—	—	—	9 (21)
Good	—	—	—	6 (14)
Average	—	—	—	7 (17)
Poor	—	—	—	2 (5)
Did not use	—	—	—	6 (14)
Did not report	—	—	—	4 (10)

[a] Percentages may not sum to 100% due to rounding.

[b] Web analytics data were available for 154 (33, 38, 41, and 42 for levels 1, 2, 3, and 4, respectively) of 161 men exposed to the intervention. The acceptability survey was sent after the 3-month intervention period; because of technical errors, 44 men received the acceptability survey late.

[c] Empty cells denote intervention features that were not offered to a given level; participants were not asked to provide feedback on features they did not receive.

A higher proportion of men (10/42, 24%) in level 4 reported that they were very satisfied with the intervention than men in levels 1, 2, and 3 (4/39, 10%; 5/42, 12%; and 5/44, 11%, respectively). A total of 6 men in levels 1 to 3 (1 in level 1, 4 in level 2, and 1 in level 3) who completed surveys at 3 months had not accessed the study website intervention (Figure 1), which may in part reflect lower satisfaction scores for these levels compared with level 4. In addition, a higher proportion of men in level 1 reported that they were dissatisfied or very dissatisfied with the intervention (5/39, 13%) compared with levels 2, 3, and 4 (1/42, 2%; 3/44, 7%; and 2/42, 5%, respectively). Men in level 4 (n=42) were also more likely to rate the intervention features as excellent or very good: exercise prescription (20/42, 48%), diet prescription (24/42, 57%), and weekly progress report (24/42, 57%) compared with the 42 men in level 2 (12/42, 29%; 13/42, 31%; 17/42, 40%, respectively), and 44 men in level 3 (9/44, 20%; 17/44, 39%; 15/44, 34%). In contrast to our expectations, a large number of men in levels 3 and 4 reported that they did not like (8/44, 18% and 4/42, 10%, respectively) or did not use (7/44, 16% and 12/42, 29%, respectively) the Fitbit. Approximately one-fourth of the men in levels 3 and 4 reported that they did not use the text messages (10/44, 23% in level 3 and 11/42, 26% in level 4). Of the 42 men in level 4, 26 (62%) completed an exercise coaching call (8 were unable to be contacted and 8 declined) and 35 (83%) completed a diet coaching call (2 were unable to be contacted and 5 declined). Of the men who received the calls, 88% (23/26) rated the exercise call as good to excellent and 69% (24/35) rated the diet call as good to excellent.

Participants' open-ended feedback on the intervention is provided in Multimedia Appendix 2. Participants noted challenges with the onboarding process, insufficient personalization of the study materials, and limited information tailored to their level and readiness for change. Although the intervention provided personalized diet and exercise prescriptions based on baseline surveys (levels 2-4), feedback from the open-ended comments indicated that several participants desired more interactive feedback, direction, and reminders. As noted above, many men indicated that they did not like or use the Fitbit; some participants already owned other devices and would have preferred to have the option to integrate those into the study website.

Lifestyle Behavior

Figure 2 shows the between-group difference in mean change (95% CI) from enrollment to 3 months comparing with the 4 levels for the overall lifestyle score. Changes between levels 2, 3, and 4 versus level 1 for the diet and physical activity scores and their subcomponents are shown in Figure 3 and Figure 4. The between-group differences in mean change, comparing each level to level 1, were as follows: 0.93 (95% CI 0.44-1.41) points for level 2; 0.51 (95% CI 0.02-0.99) points for level 3; and 1.11 (95% CI 0.65-1.57) points for level 4. The median (IQR) values for the lifestyle score and its components, by study arm and time point, are shown in Multimedia Appendix 3 and the between-group differences in mean change from enrollment to 6 months are shown in Multimedia Appendices 4-6. The small improvement in score for level 4 was attenuated, but still present, at 6 months (between-group difference in mean change level 4 vs 1: 0.72; 95% CI 0.26-1.18 points).

Level 4 had a greater improvement in diet scores at 3 months compared with level 1 (Figure 3). The between-group difference in mean change from enrollment to month 3 in level 4 as compared with level 1 was 0.49 (IQR 0.22-0.77) points. This change seemed to be driven by an increased intake of cruciferous vegetables and healthy vegetable fat and decreased intake of processed meat. As with the overall lifestyle score, the improvement was present, but attenuated, at 6 months (between-group difference in mean change in level 4 vs level 1: 0.32 [IQR 0.09-0.54]; Multimedia Appendix 5).

Similarly, for physical activity, only level 4 appeared to have a greater increase in physical activity at 3 months compared with level 1 (between-group difference in mean: 0.39; 95% CI 0.08-0.69 points; Figure 4). The change appeared to be due to small increases in aerobic exercise, strength training, and stretching. However, the difference in mean change for the overall score was not maintained at 6 months (Multimedia Appendix 6).

Figure 2. Difference in mean change in an overall lifestyle score (range 0-20) from baseline to 3 months, comparing intervention levels 2 to 4 with level 1. Higher scores indicate more healthy lifestyle behaviors. These secondary analyses included the 146 men with complete data on lifestyle behaviors at 0 and 3 months.

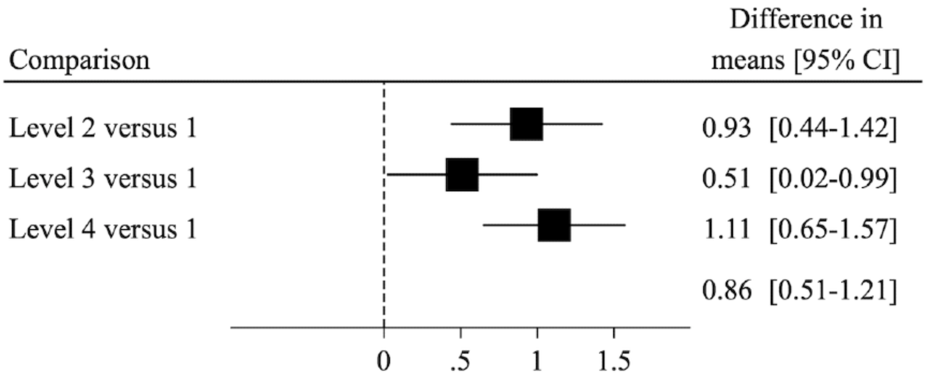


Figure 3. Differences in mean change in an overall diet score (range 0-14) and servings per day of diet items from baseline to 3 months, comparing intervention levels 2 to 4 with level 1. A higher score for the overall diet score indicates healthier diet behaviors. The intervention aimed to increase intake of cruciferous vegetables, fish, tomatoes, and vegetable fat and decrease intake of poultry with skin, processed meat, and whole milk. These secondary analyses included the 146 men with complete data on diet behaviors at 0 and 3 months.

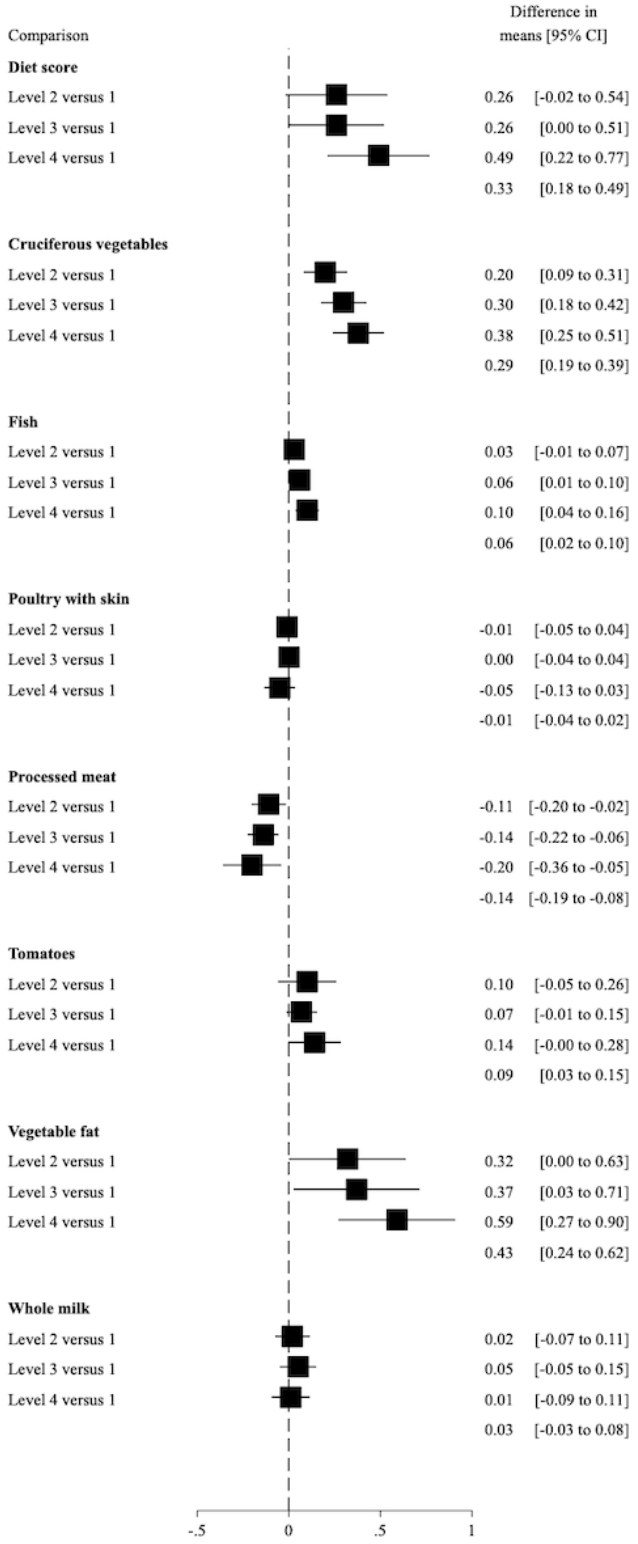
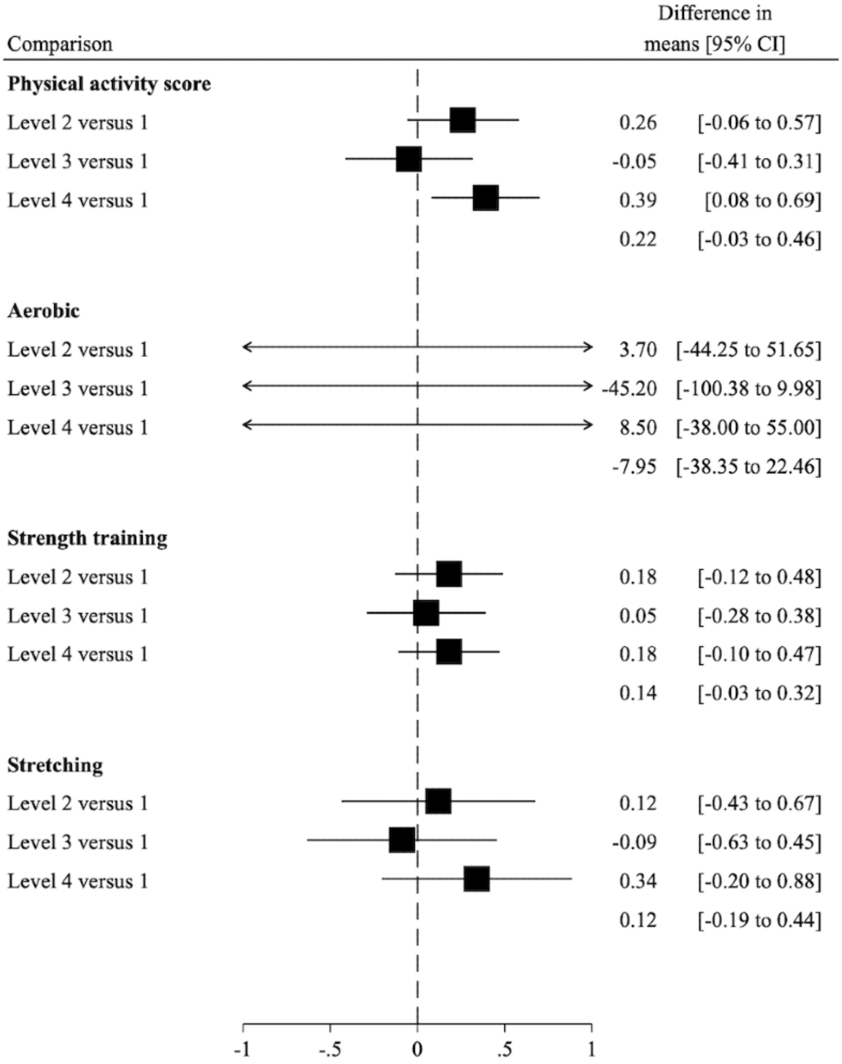


Figure 4. Difference in mean change in an overall physical activity score (range 0-6) and types of physical activity (minutes per week of aerobic; sessions per week of strength training and stretching) from baseline to 3 months, comparing intervention levels 2 to 4 with level 1. Higher scores indicate more physical activity. These secondary analyses were performed among the 152 men with complete data on physical activity at 0 and 3 months.



In a *posthoc* sensitivity analysis, we stratified men according to their self-reported aerobic activity level at the time of enrollment. As shown in Table 3, in levels 3 and 4, men who reported <90 min/wk of aerobic physical activity at enrollment appeared to report more aerobic physical activity at 3 months (median change from baseline to 3 months in level 3: 60 min/wk; IQR 30-75; level 4: 75 min/wk; IQR 30-150). Although these changes were not fully maintained, men in levels 3 and 4 who started with <90 min/wk of aerobic activity reported a median increase of 30 min/wk at 6 months compared with enrollment (level 3 IQR 0-135; level 4 IQR 0-330).

Table 3. Moderate-to-vigorous intensity aerobic physical activity minutes per week at baseline, 3 months, and 6 months among men with prostate cancer participating in a technology-supported behavioral intervention, overall and by level randomized and baseline activity.

Time point[a]	Level 1		Level 2		Level 3		Level 4	
	n	Median (IQR)	n	Median (IQR)	n	Median (IQR)	n	Median (IQR)
≥150 min/week at enrollment								
Baseline	30	345 (225 to 480)	22	405 (225 to 540)	33	345 (240 to 510)	28	398 (255 to 480)
3 months	22	263 (210 to 495)	16	323 (173 to 495)	28	285 (105 to 450)	21	435 (330 to 540)
6 months	22	398 (105 to 540)	15	345 (135 to 570)	26	383 (225 to 465)	20	338 (315 to 405)
Change, baseline to 3 months	22	−75 (−180 to 90)	16	15 (−263 to 128)	28	−53 (−360 to 75)	21	−75 (−195 to 180)
Change, baseline to 6 months	22	−38 (−150 to 120)	15	0 (−150 to 315)	26	−8 (−120 to 75)	20	−165 (−285 to 120)
90-<150 min/week at enrollment								
Baseline	6	105 (105 to 105)	9	105 (105 to 105)	3	120 (105 to 135)	10	105 (105 to 105)
3 months	4	398 (203 to 570)	8	120 (68 to 188)	3	225 (105 to 330)	8	150 (83 to 233)
6 months	4	180 (83 to 233)	8	180 (45 to 315)	3	105 (0 to 390)	8	173 (98 to 225)
Change, baseline to 3 months	4	293 (98 to 465)	8	15 (−38 to 90)	3	105 (−30 to 225)	8	45 (−23 to 120)
Change, baseline to 6 months	4	75 (−23 to 128)	8	75 (−60 to 210)	3	0 (−135 to 270)	8	68 (0 to 120)
<90 min/week at enrollment								
Baseline	13	30 (0 to 30)	18	15 (0 to 30)	14	0 (0 to 30)	13	30 (0 to 30)
3 months	10	45 (0 to 135)	14	30 (0 to 225)	9	105 (30 to 105)	9	105 (30 to 210)
6 months	10	98 (0 to 105)	13	60 (0 to 105)	7	30 (0 to 135)	9	60 (0 to 360)
Change, baseline to 3 months	10	0 (0 to 105)	14	30 (0 to 195)	9	60 (30 to 75)	9	75 (30 to 150)
Change, baseline to 6 months	10	53 (0 to 75)	13	0 (0 to 105)	7	30 (0 to 135)	9	30 (0 to 330)

[a] Two men in level 2 and 1 man in level 4 had unknown baseline physical activity.

Nonserious Adverse Events

No serious AEs occurred during the study. However, total nonserious AEs (study related and unrelated) were common in the older population of men with prostate cancer (Table 4). Most of the AEs (246/356, 69.1%) were associated with a pre-existing condition; only 15.4% (55/356) of the AEs were self-reported by the participants to be related to the study and included exacerbations of pre-existing problems. Muscle pain/injury, fatigue, and joint or bone pain were the most frequently reported, accounting for 82% (45/55) of study-related AEs. All study-related AEs were mild to moderate in severity. Although the number of study-related AEs was low, they did appear to increase across levels, particularly going from level 1 (8 study-related AEs) to level 2 (14 study-related AEs). The difference in study-related AEs across levels was driven by higher reports of muscle pain/injury and fatigue in higher levels. For example, there were 7 reports of muscle pain/injury that were deemed study related by participants in levels 3 and 4, 6 in level 2, and 3 in level 1. Study-related fatigue was highest in level 4, reported 6 times, compared with 2 reports of fatigue in level 1 and 1 report of fatigue in levels 2 and 3 each.

Table 4. Nonserious adverse events self-reported at 3 months among men with prostate cancer participating in a technology-supported behavioral intervention by randomized level.

Nonserious AEs[a]	Level 1	Level 2	Level 3	Level 4	Total
Total AE, n	80	133	76	67	356
AE related to pre-existing conditions, n	53	91	54	48	246
Study-related AE, n	8	14	15	18	55
Specific AEs, n					
. **Joint or bone pain**					
Any	21	40	24	20	105
Pre-existing	16	31	20	15	82
Study related	1	5	5	1	12
Muscle pain or injury					
Any	20	29	24	15	88
Pre-existing	9	17	14	9	49
Study related	3	6	7	7	23
Gastrointestinal issues					
Any	14	15	8	11	48
Pre-existing	11	12	5	9	37
Study related	2	1	2	3	8
Fatigue					
Any	14	24	10	16	64
Pre-existing	9	17	10	11	47
Study related	2	1	1	6	10
Dizziness or vertigo					
Any	9	15	8	3	35
Pre-existing	6	9	3	2	20
Study related	0	1	0	0	1
Shortness of breath					
Any	1	8	2	2	13
Pre-existing	1	4	2	2	9
Study related	0	0	0	1	1
Cardiovascular event					
Any	1	2	0	0	3
Pre-existing	1	1	0	0	2
Study related	0	0	0	0	0

[a] AE: adverse event.

Discussion

Principal Findings

The TrueNTH Community of Wellness study primarily sought to evaluate whether persons living with a diagnosis of prostate cancer would engage with a web-based intervention focused on diet and physical activity and secondarily explored whether such an intervention would help people adopt healthier habits. This national, multi-site, pilot randomized controlled trial demonstrated the feasibility of a technology-enhanced, remotely delivered behavioral intervention and provides insights into the acceptability of different intervention components. Overall, we met our a priori goals to enroll 200 participants in 1 year (202 men were randomized in 13 months) and retain 80% of participants at 3 months and 64% at 6 months (retention was 167/202, 82.7% and 156/202, 77.2% at 3 and 6 months, respectively).

Key takeaways from this study include the importance of an easy *onboarding* process and the value of at least some contact from a person. A high proportion of men reported difficulty registering on the portal website, and a quarter of the men failed to complete the process and did not end up receiving their assigned intervention. This underscores the need for additional orientation, email reminders, and/or follow-up calls to ensure that participants have sufficient technical support to access web-based resources. Future studies need to address how to collect sufficient personal information to deliver a tailored intervention without overburdening participants and preventing them from continuing in the program. In addition, individuals who received two 30-minute coach calls (level 4) were more satisfied with the intervention compared with other groups and appeared to be more successful in making small lifestyle changes. These results suggest that some level of coaching or human interaction is important for participant satisfaction with remotely delivered lifestyle programs. This is in agreement with previous findings [45-47]. Additional research is needed to assess whether other technology-based interactions (eg, tailored text messages or chatbots) or web-based peer-to-peer interactions can facilitate similar satisfaction and improve behavior change outcomes as coaching calls [47-50]. Overall, future studies are challenged to identify the minimum dose of health coaching needed for participant satisfaction and meaningful behavior change while maintaining scalability. These studies should consider innovative study designs that efficiently support testing multiple intervention components (eg, multiphase optimization strategy framework) [51,52].

In secondary analyses, it appeared that the intervention had small effects on lifestyle behaviors. For example, we observed a median increase of 0.5 servings per day (IQR 0.2-0.9) in cruciferous vegetable consumption in level 4 at 3 months. The changes we observed are of similar magnitude to those reported by other studies with more intensive health coaching. The Reach Out to Enhance Wellness trial conducted among long-term breast, prostate, and colorectal cancer survivors included 15 health coaching sessions over 12 months and observed an increase in *fruit and vegetable* intake of 1.1 servings per day (95% CI 0.76-1.47) when comparing the intervention group with the control group at 12 months [53]. Our results comparing the different intervention levels add to the literature and suggest that at least some *higher touch* coaching may be needed to successfully modify dietary intake.

Most men in the study did not increase their physical activity from the time of enrollment. Our study population, however, reported high levels of physical activity at enrollment and thus did not have much room for improvement. Indeed, when we stratified men based on whether they met the recommended amounts of physical activity at enrollment, we observed an increase in aerobic physical activity at 3 months among men not meeting the physical activity guidelines at baseline in levels 3 and 4. For example, the median (IQR) min/wk of moderate-to-vigorous physical activity at enrollment and 3 months among inactive men in level 4 was 30 (IQR 0-30) and 105 (IQR 30-210). This change was not maintained at 6 months. Among those who had already met the recommended 150 minutes of exercise per week at the time of enrollment, there was no increase in exercise with increasing intervention levels, and among those assigned to levels 3 and 4, there appeared to be a decrease in total minutes of exercise. It is possible that very active men altered their behavior once they realized that they exceeded the guidelines. Future studies using adaptive trial designs, such as sequential multiple assignment randomized trials, to target intervention resources to participants who need them most would be of interest.

This trial was designed to evaluate the feasibility of direct-to-patient enrollment and acceptability of a remotely delivered, web-based behavioral intervention in a study population with a wide distribution of geography and clinical disease features. Our team previously reported the results of the Prostate 8-I (P8-I) pilot study conducted at UCSF. P8-I reported a larger improvement in diet than that observed in the present trial. One difference in study design that may have played a role in the different results could be the in-clinic recruitment and on-site study visits in P8-I. Given the difference in acceptability reported by men in level 4 compared with levels 1 to 3 in this study, it is possible that having a personal connection to the study helped motivate participants in P8-I to make larger dietary changes. In addition, P8-I excluded participants who had already met 4 of 8 prespecified lifestyle recommendations, whereas this pilot did not.

Limitations

There are several limitations of this study to consider and improve upon in future trials. First, individuals who volunteered for this study were predominantly White and highly educated; thus, our results may not be generalizable to all men with prostate cancer. Further work is warranted to assess whether a remotely delivered, web-based intervention is acceptable or beneficial for more diverse or underserved populations. Most participants also had localized disease; the feasibility and acceptability of lifestyle interventions in men with more advanced disease remains to be determined. Second, diet and physical activity were assessed using self-reporting, and the instruments used may not have been sufficiently discriminative to detect small changes in lifestyle behaviors. Third, we did not include the participants' caregivers or family/friends in the intervention. In our past trial of a structured partnered exercise program for prostate cancer patients and their spouses, retention and adherence rates to exercise in patients exceeded that in our other patient-only trials [9]. Therefore, partner support may be a key facilitator for patient behavior change. Improving the health of partners may also have a positive impact on patients' health, a concept we are currently testing in a clinical exercise trial for cancer survivors and their spouse (or partners; NCT03630354). Fourth, we selected one type of physical activity tracker for integration with the web portal, and several comments made in the exit survey indicated that people would prefer more options and to use the devices they already owned (see Multimedia Appendix 2 for participant feedback). Participant feedback also indicated that additional programming to personalize and update the recommendations and messaging over time was desired. Accordingly, our team is currently enrolling individuals going to surgery for prostate cancer (Prostate 8-II trial, NCT #03999151) to a longer intervention (24 months) with more tailored feedback that adjusts to real-time self-reported diet and exercise data. P8-II also provides quarterly health coaching. Finally, it is worth noting that nonserious study-related AEs increased across intervention levels. However, the AEs reported were consistent with expected side effects of increased physical activity (eg, muscle pain, fatigue) and/or dietary change (eg, fatigue) and all were mild to moderate.

Conclusions

The TrueNTH Community of Wellness trial demonstrated the feasibility of a web-based, remotely delivered, tailored behavioral intervention among individuals with all stages of prostate cancer. Men in level 4 who received two 30-minute phone calls reported higher satisfaction, engaged more frequently with the intervention, and reported small improvements in diet and physical activity compared with men in level 1. Future studies are warranted to evaluate how to increase the effect of the intervention on lifestyle behaviors, while maintaining long-term sustainability and scalability, as well as designing and implementing behavioral interventions for more diverse populations.

Acknowledgments

The authors would like to acknowledge all the participants and Esther Moe, Carol-Ann Mullin, and Jade Moon for their assistance in managing and implementing the study. The authors also thank the Movember Foundation for supporting this project as part of its global TrueNTH initiative. This project was supported by the National Center for Advancing Translational Sciences, National Institutes of Health (NIH), Grant Numbers UL1 TR001872 and UL1TR002369, and other National Institutes of Health grants P30DK098722, P30 CA069533-21, and NCI K07CA197077. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The project was also supported by the Steven and Christine Burd-Safeway Distinguished Professorship (JC) and the Helen Diller Family Chair in Population Science for Urologic Cancer (SK).

Conflicts of Interest

SK is associated with Fellow Health Inc. TB is associated with AbbVie, Alliance Foundation Trials, Arvinas Inc, Astellas, AstraZeneca, Bayer, Boehringer Ingelheim, Clovis Oncology, Corcept Therapeutics, Endocyte Inc, GlaxoSmithKline, Janssen Biotech, Janssen Japan, Janssen Research & Development, Medivation, Inc, Merck, OncoGenex, Pfizer, Salarius Pharmaceuticals, Sotio, Theraclone Sciences/OncoResponse. JC's spouse is employed by GRAIL Inc. KP is employed by Movember. The authors are the developers/sponsors of the intervention evaluated in this report.

Multimedia Appendix 1
The components and points assigned for the lifestyle behavior score.
[DOCX File , 25 KB - jmir_v22i12e19238_app1.docx]

Multimedia Appendix 2
Participant feedback from a 3-month technology-supported behavioral intervention for prostate cancer survivors.
[DOCX File , 27 KB - jmir_v22i12e19238_app2.docx]

Multimedia Appendix 3
Baseline, 3-month, 6-month, and change from baseline lifestyle, diet, and physical activity scores among men with prostate cancer participating in a 3-month technology-supported behavioral intervention, randomized by level.
[DOCX File , 32 KB - jmir_v22i12e19238_app3.docx]

Multimedia Appendix 4
Difference in mean change in an overall lifestyle score (range 0-20) from baseline to 6 months, comparing intervention levels 2 to 4 with level 1. Higher scores indicate more healthy lifestyle behaviors.
[PNG File , 116 KB - jmir_v22i12e19238_app4.png]

Multimedia Appendix 5
Differences in mean change in an overall diet score (range 0-14) and servings per day of diet items from baseline to 6 months, comparing intervention levels 2 to 4 with level 1. A higher score for the overall diet score indicates healthier diet behaviors. The intervention aimed to increase intake of cruciferous vegetables, fish, tomatoes, and vegetable fat and decrease intake of poultry with skin, processed meat, and whole milk.
[PNG File , 171 KB - jmir_v22i12e19238_app5.png]

Multimedia Appendix 6
Difference in mean change in an overall physical activity score (range 0-6) and types of physical activity (minutes per week of aerobic; sessions per week of strength training and stretching) from baseline to 6 months, comparing intervention levels 2 to 4 with level 1. Higher scores indicate more physical activity.

[PNG File , 452 KB - jmir_v22i12e19238_app6.png]

Multimedia Appendix 7
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 1221 KB - jmir_v22i12e19238_app7.pdf]

References

1. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019 Sep;69(5):363-385 [FREE Full text] [doi: 10.3322/caac.21565] [Medline: 31184787]
2. Barry MJ, Simmons LH. Prevention of prostate cancer morbidity and mortality: primary prevention and early detection. Med Clin North Am 2017 Jul;101(4):787-806. [doi: 10.1016/j.mcna.2017.03.009] [Medline: 28577627]
3. Winters-Stone KM, Moe E, Graff JN, Dieckmann NF, Stoyles S, Borsch C, et al. Falls and frailty in prostate cancer survivors: current, past, and never users of androgen deprivation therapy. J Am Geriatr Soc 2017 Jul;65(7):1414-1419 [FREE Full text] [doi: 10.1111/jgs.14795] [Medline: 28263373]
4. Galvão DA, Taaffe DR, Spry N, Joseph D, Turner D, Newton RU. Reduced muscle strength and functional performance in men with prostate cancer undergoing androgen suppression: a comprehensive cross-sectional investigation. Prostate Cancer Prostatic Dis 2009;12(2):198-203. [doi: 10.1038/pcan.2008.51] [Medline: 18852703]
5. Chang D, Joseph DJ, Ebert MA, Galvão DA, Taaffe DR, Denham JW, et al. Effect of androgen deprivation therapy on muscle attenuation in men with prostate cancer. J Med Imaging Radiat Oncol 2014 Apr;58(2):223-228. [doi: 10.1111/1754-9485.12124] [Medline: 24118850]
6. Iadeluca L, Mardekian J, Chander P, Hopps M, Makinson GT. The burden of selected cancers in the US: health behaviors and health care resource utilization. Cancer Manag Res 2017;9:721-730 [FREE Full text] [doi: 10.2147/CMAR.S143148] [Medline: 29238222]
7. Finlay A, Wittert G, Short CE. A systematic review of physical activity-based behaviour change interventions reaching men with prostate cancer. J Cancer Surviv 2018 Aug;12(4):571-591. [doi: 10.1007/s11764-018-0694-8] [Medline: 29770953]
8. Winters-Stone KM, Dobek JC, Bennett JA, Dieckmann NF, Maddalozzo GF, Ryan CW, et al. Resistance training reduces disability in prostate cancer survivors on androgen deprivation therapy: evidence from a randomized controlled trial. Arch Phys Med Rehabil 2015 Jan;96(1):7-14 [FREE Full text] [doi: 10.1016/j.apmr.2014.08.010] [Medline: 25194450]
9. Winters-Stone KM, Lyons KS, Dobek J, Dieckmann NF, Bennett JA, Nail L, et al. Benefits of partnered strength training for prostate cancer survivors and spouses: results from a randomized controlled trial of the Exercising Together project. J Cancer Surviv 2016 Aug;10(4):633-644. [doi: 10.1007/s11764-015-0509-0] [Medline: 26715587]
10. Fornusek CP, Kilbreath SL. Exercise for improving bone health in women treated for stages I-III breast cancer: a systematic review and meta-analyses. J Cancer Surviv 2017 Oct;11(5):525-541. [doi: 10.1007/s11764-017-0622-3] [Medline: 28639157]
11. Winters-Stone KM, Dobek JC, Bennett JA, Maddalozzo GF, Ryan CW, Beer TM. Skeletal response to resistance and impact training in prostate cancer survivors. Med Sci Sports Exerc 2014 Aug;46(8):1482-1488 [FREE Full text] [doi: 10.1249/MSS.0000000000000265] [Medline: 24500540]
12. Richman EL, Kenfield SA, Stampfer MJ, Paciorek A, Carroll PR, Chan JM. Physical activity after diagnosis and risk of prostate cancer progression: data from the cancer of the prostate strategic urologic research endeavor. Cancer Res 2011 Jun 1;71(11):3889-3895 [FREE Full text] [doi: 10.1158/0008-5472.CAN-10-3932] [Medline: 21610110]
13. Kenfield SA, Stampfer MJ, Giovannucci E, Chan JM. Physical activity and survival after prostate cancer diagnosis in the health professionals follow-up study. J Clin Oncol 2011 Feb 20;29(6):726-732 [FREE Full text] [doi: 10.1200/JCO.2010.31.5226] [Medline: 21205749]
14. Friedenreich CM, Wang Q, Neilson HK, Kopciuk KA, McGregor SE, Courneya KS. Physical activity and survival after prostate cancer. Eur Urol 2016 Oct;70(4):576-585 [FREE Full text] [doi: 10.1016/j.eururo.2015.12.032] [Medline: 26774959]
15. Bonn SE, Sjölander A, Lagerros YT, Wiklund F, Stattin P, Holmberg E, et al. Physical activity and survival among men diagnosed with prostate cancer. Cancer Epidemiol Biomarkers Prev 2015 Jan;24(1):57-64 [FREE Full text] [doi: 10.1158/1055-9965.EPI-14-0707] [Medline: 25527697]
16. Wang Y, Jacobs EJ, Gapstur SM, Maliniak ML, Gansler T, McCullough ML, et al. Recreational physical activity in relation to prostate cancer-specific mortality among men with nonmetastatic prostate cancer. Eur Urol 2017 Dec;72(6):931-939. [doi: 10.1016/j.eururo.2017.06.037] [Medline: 28711382]
17. Yang M, Kenfield SA, van Blarigan EL, Batista JL, Sesso HD, Ma J, et al. Dietary patterns after prostate cancer diagnosis in relation to disease-specific and total mortality. Cancer Prev Res (Phila) 2015 Jun;8(6):545-551 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-14-0442] [Medline: 26031631]
18. van Blarigan EL, Kenfield SA, Yang M, Sesso HD, Ma J, Stampfer MJ, et al. Fat intake after prostate cancer diagnosis and mortality in the Physicians' Health Study. Cancer Causes Control 2015 Aug;26(8):1117-1126 [FREE Full text] [doi: 10.1007/s10552-015-0606-4] [Medline: 26047644]

19. Richman EL, Stampfer MJ, Paciorek A, Broering JM, Carroll PR, Chan JM. Intakes of meat, fish, poultry, and eggs and risk of prostate cancer progression. Am J Clin Nutr 2010 Mar;91(3):712-721 [FREE Full text] [doi: 10.3945/ajcn.2009.28474] [Medline: 20042525]
20. Richman EL, Carroll PR, Chan JM. Vegetable and fruit intake after diagnosis and risk of prostate cancer progression. Int J Cancer 2012 Jul 1;131(1):201-210 [FREE Full text] [doi: 10.1002/ijc.26348] [Medline: 21823116]
21. Richman EL, Kenfield SA, Stampfer MJ, Giovannucci EL, Chan JM. Egg, red meat, and poultry intake and risk of lethal prostate cancer in the prostate-specific antigen-era: incidence and survival. Cancer Prev Res (Phila) 2011 Dec;4(12):2110-2121 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-11-0354] [Medline: 21930800]
22. Peisch SF, van Blarigan EL, Chan JM, Stampfer MJ, Kenfield SA. Prostate cancer progression and mortality: a review of diet and lifestyle factors. World J Urol 2017 Jun;35(6):867-874 [FREE Full text] [doi: 10.1007/s00345-016-1914-3] [Medline: 27518576]
23. Pettersson A, Kasperzyk JL, Kenfield SA, Richman EL, Chan JM, Willett WC, et al. Milk and dairy consumption among men with prostate cancer and risk of metastases and prostate cancer death. Cancer Epidemiol Biomarkers Prev 2012 Mar;21(3):428-436 [FREE Full text] [doi: 10.1158/1055-9965.EPI-11-1004] [Medline: 22315365]
24. Chavarro JE, Stampfer MJ, Hall MN, Sesso HD, Ma J. A 22-y prospective study of fish intake in relation to prostate cancer incidence and mortality. Am J Clin Nutr 2008 Nov;88(5):1297-1303 [FREE Full text] [doi: 10.3945/ajcn.2008.26419] [Medline: 18996866]
25. Chan JM, Holick CN, Leitzmann MF, Rimm EB, Willett WC, Stampfer MJ, et al. Diet after diagnosis and the risk of prostate cancer progression, recurrence, and death (United States). Cancer Causes Control 2006 Mar;17(2):199-208. [doi: 10.1007/s10552-005-0413-4] [Medline: 16425098]
26. Dickerman BA, Giovannucci E, Pernar CH, Mucci LA, Hernán MA. Guideline-based physical activity and survival among US men with nonmetastatic prostate cancer. Am J Epidemiol 2019 Mar 1;188(3):579-586 [FREE Full text] [doi: 10.1093/aje/kwy261] [Medline: 30496346]
27. Rock CL, Doyle C, Demark-Wahnefried W, Meyerhardt J, Courneya KS, Schwartz AL, et al. Nutrition and physical activity guidelines for cancer survivors. CA Cancer J Clin 2012;62(4):243-274 [FREE Full text] [doi: 10.3322/caac.21142] [Medline: 22539238]
28. van der Bij AK, Laurant MG, Wensing M. Effectiveness of physical activity interventions for older adults: a review. Am J Prev Med 2002 Feb;22(2):120-133. [doi: 10.1016/s0749-3797(01)00413-5] [Medline: 11818183]
29. Demark-Wahnefried W, Morey MC, Sloane R, Snyder DC, Miller PE, Hartman TJ, et al. Reach out to enhance wellness home-based diet-exercise intervention promotes reproducible and sustainable long-term improvements in health behaviors, body weight, and physical functioning in older, overweight/obese cancer survivors. J Clin Oncol 2012 Jul 1;30(19):2354-2361 [FREE Full text] [doi: 10.1200/JCO.2011.40.0895] [Medline: 22614994]
30. Keogh JWL, Patel A, MacLeod RD, Masters J. Perceived barriers and facilitators to physical activity in men with prostate cancer: possible influence of androgen deprivation therapy. Eur J Cancer Care (Engl) 2014 Mar;23(2):263-273. [doi: 10.1111/ecc.12141] [Medline: 24134506]
31. Kuijpers W, Groen WG, Aaronson NK, van Harten WI. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013 Feb 20;15(2):e37 [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]
32. Golsteijn RH, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Development of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: OncoActive. BMC Cancer 2017 Jun 26;17(1):446 [FREE Full text] [doi: 10.1186/s12885-017-3397-z] [Medline: 28651586]
33. Golsteijn RH, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Short-term efficacy of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2018 Oct 30;15(1):106 [FREE Full text] [doi: 10.1186/s12966-018-0734-9] [Medline: 30376857]
34. Mohamad H, Ntessalen M, Craig LC, Clark J, Fielding S, N'Dow J, et al. A self-help diet and physical activity intervention with dietetic support for weight management in men treated for prostate cancer: pilot study of the Prostate Cancer Weight Management (PRO-MAN) randomised controlled trial. Br J Nutr 2019 Sep 14;122(5):592-600. [doi: 10.1017/S0007114519001090] [Medline: 31177994]
35. Trinh L, Arbour-Nicitopoulos KP, Sabiston CM, Berry SR, Loblaw A, Alibhai SM, et al. RiseTx: testing the feasibility of a web application for reducing sedentary behavior among prostate cancer survivors receiving androgen deprivation therapy. Int J Behav Nutr Phys Act 2018 Jun 7;15(1):49 [FREE Full text] [doi: 10.1186/s12966-018-0686-0] [Medline: 29880049]
36. van de Wiel HJ, Stuiver MM, May AM, van Grinsven S, Aaronson NK, Retèl VP, et al. (Cost-)effectiveness of an internet-based physical activity support program (with and without physiotherapy counselling) on physical activity levels of breast and prostate cancer survivors: design of the PABLO trial. BMC Cancer 2018 Nov 6;18(1):1073 [FREE Full text] [doi: 10.1186/s12885-018-4927-z] [Medline: 30400784]
37. Mehra S, Visser B, Dadema T, van den Helder J, Engelbert RH, Weijs PJ, et al. Translating behavior change principles into a blended exercise intervention for older adults: design study. JMIR Res Protoc 2018 May 2;7(5):e117 [FREE Full text] [doi: 10.2196/resprot.9244] [Medline: 29720358]

38. Kloek CJ, van Dongen JM, de Bakker DH, Bossen D, Dekker J, Veenhof C. Cost-effectiveness of a blended physiotherapy intervention compared to usual physiotherapy in patients with hip and/or knee osteoarthritis: a cluster randomized controlled trial. BMC Public Health 2018 Aug 31;18(1):1082 [FREE Full text] [doi: 10.1186/s12889-018-5975-7] [Medline: 30170586]
39. Groen WG, van Harten WH, Vallance JK. Systematic review and meta-analysis of distance-based physical activity interventions for cancer survivors (2013-2018): We still haven't found what we're looking for. Cancer Treat Rev 2018 Sep;69:188-203. [doi: 10.1016/j.ctrv.2018.07.012] [Medline: 30077954]
40. Winters-Stone KM, Kenfield SA, Van Blarigan EL, Moe EL, Ramsdill JW, Daniel K, et al. Effect of increasing levels of web-based behavioral support on changes in physical activity, diet, and symptoms in men with prostate cancer: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Nov 15;7(11):e11257 [FREE Full text] [doi: 10.2196/11257] [Medline: 30442638]
41. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
42. Stewart AL, Mills KM, King AC, Haskell WL, Gillis D, Ritter PL. CHAMPS physical activity questionnaire for older adults: outcomes for interventions. Med Sci Sports Exerc 2001 Jul;33(7):1126-1141. [doi: 10.1097/00005768-200107000-00010] [Medline: 11445760]
43. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992 May 15;135(10):1114-26; discussion 1127. [doi: 10.1093/oxfordjournals.aje.a116211] [Medline: 1632423]
44. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
45. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018 Apr 11;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]
46. McTigue KM, Conroy MB, Hess R, Bryce CL, Fiorillo AB, Fischer GS, et al. Using the internet to translate an evidence-based lifestyle intervention into practice. Telemed J E Health 2009 Nov;15(9):851-858. [doi: 10.1089/tmj.2009.0036] [Medline: 19919191]
47. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
48. Bibault J, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. J Med Internet Res 2019 Nov 27;21(11):e15787 [FREE Full text] [doi: 10.2196/15787] [Medline: 31774408]
49. Greer S, Ramo D, Chang Y, Fu M, Moskowitz J, Haritatos J. Use of the chatbot 'Vivibot' to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]
50. Piau A, Crissey R, Brechemier D, Balardy L, Nourhashemi F. A smartphone Chatbot application to optimize monitoring of older patients with cancer. Int J Med Inform 2019 Aug;128:18-23. [doi: 10.1016/j.ijmedinf.2019.05.013] [Medline: 31160007]
51. Kelleher SA, Dorfman CS, Plumb Vilardaga JC, Majestic C, Winger J, Gandhi V, et al. Optimizing delivery of a behavioral pain intervention in cancer patients using a sequential multiple assignment randomized trial SMART. Contemp Clin Trials 2017 Jun;57:51-57 [FREE Full text] [doi: 10.1016/j.cct.2017.04.001] [Medline: 28408335]
52. Phillips SM, Collins LM, Penedo FJ, Courneya KS, Welch W, Cottrell A, et al. Optimization of a technology-supported physical activity intervention for breast cancer survivors: Fit2Thrive study protocol. Contemp Clin Trials 2018 Mar;66:9-19 [FREE Full text] [doi: 10.1016/j.cct.2018.01.001] [Medline: 29330081]
53. Morey MC, Snyder DC, Sloane R, Cohen HJ, Peterson B, Hartman TJ, et al. Effects of home-based diet and exercise on functional outcomes among older, overweight long-term cancer survivors: RENEW: a randomized controlled trial. J Am Med Assoc 2009 May 13;301(18):1883-1891 [FREE Full text] [doi: 10.1001/jama.2009.643] [Medline: 19436015]

Abbreviations

AE: adverse event
CHAMPS: Community Health Activities Model Program for Seniors
FFQ: food frequency questionnaire
OHSU: Oregon Health and Sciences University
NIH: National Institutes of Health
P8-I: Prostate 8-I
UCD: University of Colorado Denver

UCSF: University of California San Francisco

Edited by G Eysenbach; submitted 14.04.20; peer-reviewed by W van Harten, C Short; comments to author 13.06.20; revised version received 07.08.20; accepted 11.11.20; published 31.12.20.

Please cite as:
Chan JM, Van Blarigan EL, Langlais CS, Zhao S, Ramsdill JW, Daniel K, Macaire G, Wang E, Paich K, Kessler ER, Beer TM, Lyons KS, Broering JM, Carroll PR, Kenfield SA, Winters-Stone KM
Feasibility and Acceptability of a Remotely Delivered, Web-Based Behavioral Intervention for Men With Prostate Cancer: Four-Arm Randomized Controlled Pilot Trial
J Med Internet Res 2020;22(12):e19238
URL: http://www.jmir.org/2020/12/e19238/
doi:10.2196/19238
PMID:33382378

Original Paper

Internet-Based Support and Coaching With Complementary Clinic Visits for Young People With Attention-Deficit/Hyperactivity Disorder and Autism: Controlled Feasibility Study

Helena Sehlin[1], MSc; Britt Hedman Ahlström[2], PhD; Ingrid Bertilsson[3], MSc; Gerhard Andersson[4], PhD; Elisabet Wentz[5], PhD

[1]Gillberg Neuropsychiatry Centre, Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden

[2]Department of Health Sciences, University West, Trollhättan, Sweden

[3]Habilitation & Health, Region Västra Götaland, Skövde, Sweden

[4]Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

[5]Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden

Corresponding Author:
Helena Sehlin, MSc
Gillberg Neuropsychiatry Centre
Institute of Neuroscience and Physiology
University of Gothenburg
Kungsgatan 12
Gothenburg, SE 411 19
Sweden
Phone: 46 736838212
Email: helena.sehlin@vgregion.se

Abstract

Background: Individuals with attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) can experience obstacles in traditional health care situations due to difficulties associated with their impairment.

Objective: This controlled study aims to investigate the feasibility of an internet-based support and coaching intervention (IBSC), including 2 weekly chat sessions and 2 complementary clinic visits with coaches over the course of 8 weeks, for adolescents and young adults with ADHD and/or ASD in 2 naturalistic routine care settings.

Methods: Individuals with ADHD and/or ASD aged 15-32 years were recruited in 2 clinical settings, where they received either IBSC (n=24) or treatment as usual (TAU; n=20). Outcome measures included self-report questionnaires assessing quality of life (Manchester Short Assessment for Quality of Life), sense of coherence (Sense Of Coherence 29), self-esteem (Rosenberg Self-Esteem Scale), and anxiety and depressive symptoms (Hospital Anxiety and Depression Scale [HADS] and Montgomery-Åsberg Depression Rating Scale-Self-reported, respectively).

Results: Significant between-group effects were observed in measures of anxiety (HADS) at postintervention (P=.02) as well as at the 6-month follow-up (P=.004). Significant between-group effects were also noted for depressive symptoms (HADS) postintervention (P=.04). The between-group effects were partially explained by a deterioration in the TAU group. A significant increase in self-esteem (P=.04) as well as a decrease in anxiety (P=.003) at the 6-month follow-up was observed in the intervention group following IBSC. Findings from a qualitative study of the intervention are consistent with the results.

Conclusions: The findings from this study suggest that IBSC holds promise as a feasible complement or alternative to traditional face-to-face health care meetings.

(J Med Internet Res 2020;22(12):e19658) doi:10.2196/19658

KEYWORDS

attention-deficit/hyperactivity disorder; autism; coaching; internet-based intervention; social support

Introduction

Background

In the last two decades, neurodevelopmental disorders (NDDs), such as attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD), have increasingly been recognized and diagnosed [1-4]. ADHD is estimated to affect 5% of the child and adolescent population and approximately 2.5% of the adult population [5,6]. The prevalence rates for ASD are estimated to be from 1% to 2.6% [1,7,8]. ADHD and ASD are complex and disabling conditions, which often coocur with other NDDs [2,9]. If left unmanaged, they can lead to negative consequences, including deterioration of school and work performance, relationships, living situation, self-esteem, and overall quality of life [10-12].

Increased assessment and diagnosis rates for individuals with NDDs have several long-term benefits, including increased access to early treatment for children, which could reduce the effects of core deficits and lessen the occurrence of secondary psychiatric conditions [13]. However, it also calls for efficient treatment and support for adolescents and adults with these conditions. The transition from adolescence into adulthood has been acknowledged as particularly challenging for these individuals [14,15]. This period in life often characterized by a decrease in structure and support at home and in school and increased demands to manage daily living independently. Transitioning from child to adult health care services can also result in a decrease in support for individuals with NDDs [16,17].

Current recommendations for adolescents and adults with ADHD comprise multimodal approaches, including pharmacological interventions, psychoeducation, environmental modifications, and behavioral interventions (eg, cognitive behavioral therapy [CBT]) [10,18,19]. Psychostimulants are often the first-line treatment and have been proven efficacious for core symptoms [10,18], but in approximately 50%, medication alone fails to manage the condition and/or comes with impeding side effects [18,20]. Evidence for CBT remains promising, yet preliminary [18]. Coaching is increasingly recommended as a component of multimodal treatment for adolescents and adults with ADHD [21,22]. Coaching includes individual goal setting, psychoeducation, and a collaborative approach to handle the individual's main problems. ADHD coaching also relies heavily upon the coach who should possess significant experience and a thorough understanding of the disorder [21,23,24].

Different supports have been developed and tested for ASD, but there is still a great paucity of research within this area. Guidelines promote interventions focusing on communication, interaction, and management of concomitant mental health problems through, for example, CBT [23,24]. In general, guidelines for ASD stress the necessity of making adjustments to the environment, not least to facilitate equal access to health care services [23,24]. There is a lack of studies pertaining to coaching for teens or young adults with ASD. However, a recent review concluded that there is a need for personalized support focusing on daily life problems (eg, mentoring) as opposed to narrow skills training [25].

Internet-delivered treatments have been proven effective for a number of psychiatric disorders [26]. The utilization of chat-based communication in offering psychological support has also increased over the years. A review revealed that chat-based interventions had mostly better or equivalent outcomes when compared with wait-list control and treatment as usual (TAU) [27] and the actual acceptability of this support has been shown to be high, even among people with severe mental health problems [28]. Furthermore, text- and internet-based communication is a familiar part of young people's everyday life [29].

Individuals with NDDs often struggle with traditional means of support, for example, clinic visits. Nonverbal communication such as eye gaze and gestures can be a challenge for people with ASD, and internet-based support eliminates these issues. Internet-based interventions are also flexible and accessible, which is facilitating for individuals with ADHD who often have trouble with time management and organization. Overall, communication through the internet could prove to be beneficial for individuals with NDDs [30-32]. However, very few interventions exist targeting this specific population. Some data have been published regarding internet-based CBT for ADHD [33] and internet-delivered psychoeducation for adolescents and adults with ASD [34].

Objectives

Individuals with NDDs comprise a largely heterogeneous group that can experience obstacles in traditional health care. Hence, it is of paramount importance to develop a more needs-based approach to support, especially when transitioning into adulthood. The main objective of this study is to investigate whether an internet-based support and coaching model (IBSC) can be feasible for these individuals. The intervention was first attempted in a small validation study [35] and showed an improved sense of coherence, self-esteem, and quality of life. This study aims to replicate the results and assess the feasibility of the model in 2 naturalistic clinical contexts using a larger sample and a comparison group. Our main hypotheses, based on the previous results, are that an increase in self-esteem, sense of coherence, and subjective quality of life would be observed immediately after the intervention. We also expected symptom reduction for anxiety and depressive symptoms and increased overall quality of life compared with the comparison group.

Methods

Study Design

This study (ClinicalTrials.gov Identifier NCT02300597) was designed as a nonrandomized controlled feasibility study. Participants were recruited to either the intervention (IBSC) or the TAU. They were not randomized due to previous experiences regarding recruitment, that is, individuals with ASD have difficulties tolerating uncertainty and were reluctant to participate without knowing the study conditions they were to be allocated. Outcome data in the form of self-report scales were completed before the start of the intervention, after the intervention (8 weeks after baseline for the control group), and 6 months later. Both the intervention and comparison groups received TAU between the 8-week and 6-month follow-up.

Participants and Recruitment

Participants were consecutively recruited between autumn of 2010 and autumn of 2014 at 2 study centers in the southwest of Sweden: a habilitation center providing assistance for children, adolescents, and adults with NDDs including ASD and an outpatient psychiatric clinic specialized in adults with ASD and ADHD. Participants needed to fulfill the following inclusion criteria: (1) being 15 to 32 years of age, (2) having prior confirmed diagnosis of ADHD, ASD, or both according to the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV), (3) having access to a computer with internet connection, and (4) no other ongoing support or psychological treatment during the study period (only pertaining to the intervention group). Exclusion criteria were ongoing psychosis, serious and ongoing alcohol and/or substance misuse disorder, major depressive disorder (if an obstacle to conform with the intervention or in better need of other treatment), conduct disorder/antisocial personality disorder, severe dyslexia, and known intellectual disability.

Participants were asked to take part during regular visits at the study sites and received a description of the study, including internet security. Eligibility was confirmed through an interview by a clinical psychologist or psychiatrist using the alcohol/substance use and psychotic disorder modules of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) [36] and the antisocial personality disorder module of the Structured Clinical Interview for DSM-IV Axis II Disorder (SCID II) [37]. Participants completed all outcome measures (see the Instruments section). The Montgomery Åsberg Depression Scale (MADRS) [38] was used both as an outcome measure and to screen for symptoms of depression at baseline.

A total of 31 individuals agreed to participate in the intervention. They were assessed to be eligible for participation and included in the intervention arm. Before the intervention, an individual chose to refrain, leading to a total of 97% (30/31) of individuals in the intervention group.

During the course of the study, comparison cases were recruited at each study site. They were matched by age, gender, and NDD diagnoses. In the latter part of the study, some exceptions in terms of matching were made due to recruitment problems. Comparison cases underwent the same procedure with regard to eligibility with respect to inclusion and exclusion criteria. A total of 21 comparison cases were recruited, and all but one (20/21, 95%), who was deemed too depressed, were eligible and included. Figure 1 shows the flow of participants in a CONSORT (Consolidated Standards of Reporting Trials) flowchart. Table 1 provides demographic and clinical characteristics of participants at baseline; only 80% (24/30) completed IBSC; therefore, only 24 were presented.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart of participants. Regarding TAU, after assessment for eligibility, one individual was excluded due to major depression, leaving 20 to be allocated to TAU. IBSC: internet-based support and coaching; TAU: treatment as usual.

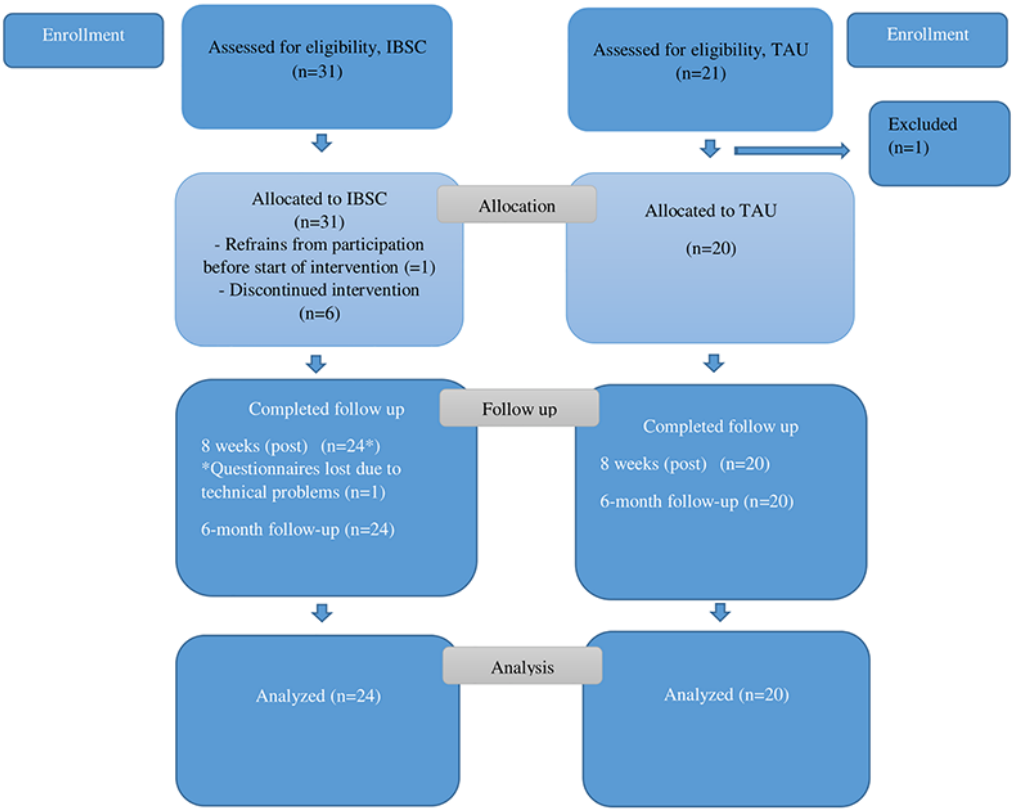


Table 1. Demographics and sample characteristics at baseline.

Demographic	Intervention (n=24[a])	Treatment as usual (n=20)	*P* value
Age at inclusion (years)			
Mean (SD)	21.0 (5.1)	22.1 (5.1)	.48
Median (range)	20 (15.0-32.0)	22.0 (15.0-32.0)	.48
Gender, n (%)			
Male	13 (54)	10 (50)	N/A[b]
Female	11 (46)	10 (50)	>.99
Diagnosis, n (%)			
ASD[c]	9 (38)	7 (35)	N/A
ADHD[d]	3 (13)	5 (25)	N/A
ASD+ADHD	12 (50)	8 (40)	.55
GAF[e] score (10-point interval)			
31-40	9 (38)	3 (15)	N/A
41-50	9 (38)	4 (20)	N/A
51-60	5 (21)	8 (40)	N/A
61-70	1 (4)	5 (25)	.006
Study center (1 and 2), n (%)			
1	11 (46)	11 (55)	N/A
2	13 (54)	9 (45)	.76
Geographical area, n (%)			
Urban	13 (54)	17 (85)	N/A
Rural	11 (46)	3 (15)	.06
Civil state, n (%)			
Married/living with partner	2 (8)	3 (15)	N/A
In a relationship (not living together)	2 (8)	6 (30)	N/A
Single	20 (83)	11 (55)	.11
Living situation, n (%)			
One-person household	8 (33)	6 (30)	N/A
Living with partner and/or children	2 (8)	5 (25)	N/A
Living with parents and/or siblings	14 (58)	8 (40)	N/A
Living with friends and/or acquaintances	0 (0)	1 (5)	.28
Level of education (completed or ongoing)			
Has not completed compulsory school[f]	2 (8)	0 (0)	N/A
Compulsory school	7 (29)	8 (42)	N/A
Upper secondary school	11 (46)	9 (47)	N/A
Vocational education (after compulsory school)	0 (0)	1 (5)	N/A
University	4 (17)	1 (5)	.65
Missing	0 (0)	1 (5)	N/A
Occupation, n (%)			
Employed	0 (0)	7 (35)	N/A
Unemployed	6 (25)	3 (15)	N/A
Student[g]	13 (54)	7 (35)	N/A

Demographic	Intervention (n=24[a])	Treatment as usual (n=20)	*P* value
Work experience placement	2 (8)	1 (5)	N/A
Sick leave	3 (13)	2 (10)	.04
Support from social services[h], n (%)			
Yes	7 (29)	7 (37)	N/A
No	17 (71)	12 (63)	.83
Missing	0 (0)	1 (5)	N/A
ADHD medication at baseline[i], n (%)			
No	14 (58)	11 (55)	N/A
Yes	10 (42)	9 (45)	>.99

[a] Only 24 out of 30 in the intervention group completed an internet-based support and coaching intervention; therefore, only 24 are presented in the table.

[b] N/A: not applicable.

[c] ASD: autism spectrum disorder.

[d] ADHD: attention-deficit/hyperactivity disorder.

[e] GAF: Global Assessment of Functioning.

[f] No formal education or terminated compulsory school without complete grades.

[g] All levels of education (eg, compulsory school, secondary school, vocational education, and university studies).

[h] Support can include assistance through the Swedish Act concerning Support and Service for Persons with Certain Functional Impairments (LSS) or from Social Services, for example, so called *contact person*, relief service, or living support.

[i] In all cases but two this was psychostimulant medication.

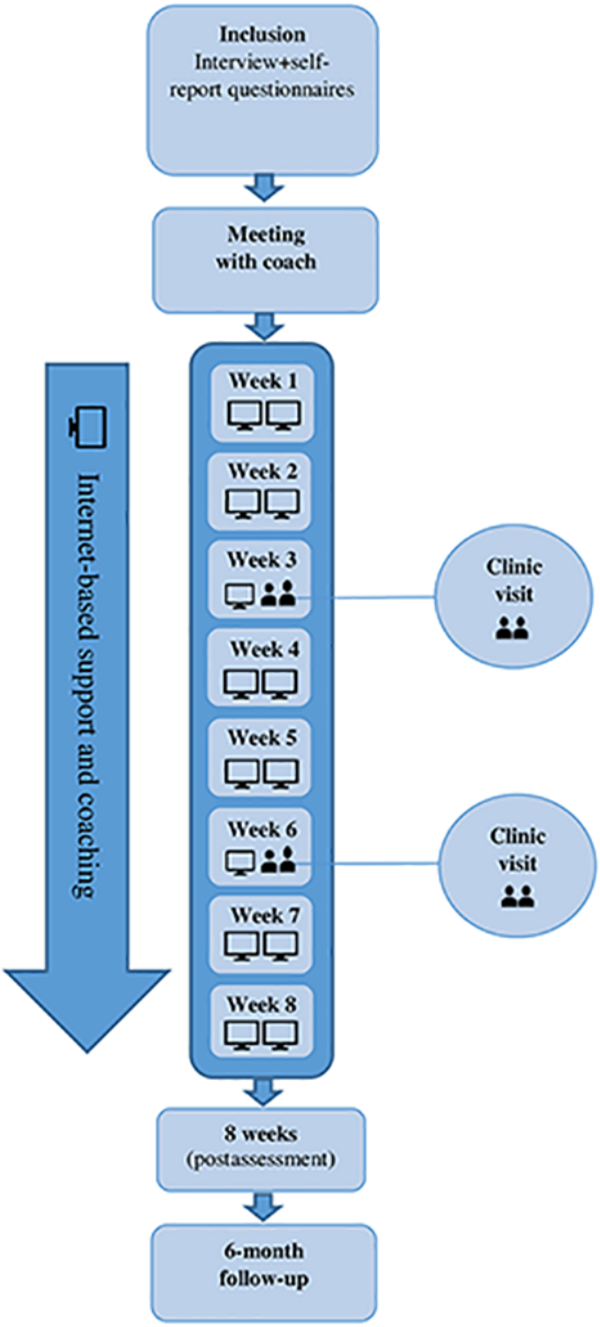
Overall, the recruitment period spanned over a relatively long period because it took place in routine care settings where the staff had restricted time for the study in relation to parallel practice of routine psychiatric care and/or habilitating interventions. In addition, the participating units were relatively small and each recruitment was immediately followed by coaches engaging in the subsequent intervention (as well as the accompanying administration).

Intervention and TAU

Intervention

The model for IBSC was developed and validated in a study by Wentz et al [35]. The model consists of 8 weeks of internet-based support and coaching mediated through a chat program. It includes 2 chat sessions in a week and 1 individual face-to-face meeting in the clinical setting (replacing chat sessions) during weeks 3 and 6 of the intervention. Each chat session lasted between 30 min and 1 hour. Short communication is also made possible through an email function (Figure 2 provides the IBSC flowchart). The intervention seeks to provide individualized psychoeducation about ADHD and/or ASD as well as support regarding aspects of daily life. The intervention started with a meeting with an appointed coach, with the purpose of becoming acquainted, to inform about the chat program and to help set it up. The coach and participant discussed preliminary areas of focus for the chat sessions as requested by the participant (ie, social interaction, daily routines, stress, study technique, time management). Support and coaching are highly individualized, based on the participants' expressed needs and requests at each chat session but also continuously readdressing the areas initially agreed on. Coaches validate and acknowledge participants in their experienced difficulties, discuss possible explanations (ie, psychoeducation), empower and encourage participants to find coping strategies, and offer specific advice on managing conveyed everyday life problems. Support and coaching are firmly based on the knowledge and clinical experience of NDDs and is intentionally more focused on everyday support as opposed to highly structured and narrow skill-specific interventions (eg, manualized treatments such as CBT). Coaches were all educated health care professionals, including occupational therapists, clinical psychologists, social workers, and special education teachers, employed at the 2 study sites. All coaches had extensive experience in ASD and ADHD. Before the start of the chat project, coaches were thoroughly informed about the aims and basis of the model, that is, support and coaching in daily life problems as opposed to more serious mental health matters that could require other kinds of help. Every second week, supervision was offered to coaches and research coordinators by the head of the project (the last author; EW). Supervision served as an opportunity to discuss issues concerning the support, model fidelity, and recruitment matters.

Figure 2. Flowchart of the internet-based support and coaching model. Assessment at 8 weeks (postassessment): assessment was carried out immediately after the end of the 8-week intervention.



TAU

TAU comprised any typically occurring treatment, that is, pharmacological treatment and/or psychological treatment, psychosocial support, occupational therapy (interventions pertaining to structure in daily living), and group psychoeducation (with most individuals receiving mainly pharmacological treatment including physical check-ups). Some individuals did not receive any active treatment during the study period (due to difficulties in complying with offered support).

Instruments

This study used the same instruments as those used in the validation study [35]. All outcome measures were prespecified according to the CONSORT protocol.

Primary Outcome Measures

Quality of life (QoL), as assessed by the Manchester Short Assessment for Quality of Life (MANSA), was used as a primary outcome measure in accordance with the validation study. The scale includes 4 objective QoL questions to be answered with *yes* or *no* and 12 subjective questions concerning job, economy, friendships, leisure activities, accommodation, personal safety, living situation, sex life, relationship with family, and physical and mental health [39]. As the study included participants as young as 15 years of age, the scale was modified to exclude the question about sex life. The remaining 11 items were assessed on a 7-point Likert scale (1=negative extreme and 7=positive extreme) and summarized into a total score. The Swedish version of MANSA has been proven to have good reliability and validity in patients with mental illness and has also shown satisfactory reliability in terms of internal consistency [40]. In line with the findings of our validation study, the effects on the first item of the MANSA assessing subjective QoL were analyzed [35].

Secondary Outcome Measures

Symptoms of depression and anxiety were measured using the Hospital Anxiety and Depression Scale (HADS) [41]. The HADS consists of 14 questions, with 7 questions constituting 2 subscales: anxiety and depression. Each item is rated on a 4-point scale. The Montgomery-Åsberg Depression Rating Scale-Self-reported (MADRS-S) was also used as a measure of depressive symptoms. It consists of 9 questions that can be scored from 0 to 6 points [38]. In both the HADS subscales as well as in the MADRS-S, a higher score represents a higher symptom burden.

The Sense of Coherence (SOC 29) scale is a life orientation questionnaire built on the concept of salutogenesis and has 3 main components: comprehensibility, manageability, and meaningfulness [42]. The SOC has 29 items, each rated on a 7-point scale where a higher total score represents a better ability to cope with stress and to stay healthy [43]. The Rosenberg Self-Esteem Scale (RSES) was used to measure self-esteem. The questionnaire comprises 10 questions each rated on a 4-point scale. A higher total score reflects better self-esteem [44,45]. Sociodemographic information was collected through a self-report questionnaire (Table 1). Background information on diagnosis, medication, and other received treatment and support was collected through medical records.

Global Assessment of Functioning

To assess the participants' general level of functioning at baseline, all individuals were evaluated retrospectively using the Global Assessment of Functioning (GAF) Scale from DSM-IV- Text Revision [46]. The scale ranged from 0 (severe impairment) to 100 (very high functioning) with descriptions for every 10-point interval considering psychiatric symptoms and social and occupational functioning. The assessment was based on current diagnoses, available sociodemographic information, level of anxiety and depression, and participant responses on the item level for the MANSA. Although not reported in this study, a questionnaire measuring perceived caregiver burden and the individuals functioning in several areas had been completed by next of kin for each participant before the start of the intervention [47]. Information from this questionnaire was similarly used at the item level to assess global functioning. All information except for gender and study arm was summarized into separate case presentations and assessed for GAF score by the main author (HS). Cases were individually and blindly assessed by the last author (EW). As the assessment was conducted retrospectively, a 10-point interval was used, which was deemed more accurate than an exact rating. Interrater reliability was assessed using Cohen kappa (k) and found to be very good (k=0.89). The GAF results are shown in Table 1.

Statistical Analysis

All statistical analyses were performed using SAS Software version 9.4 (SAS Institute Inc). Differences in baseline characteristics between the intervention and the TAU group (Table 1), as well as between the intervention group and loss to follow-up (an attrition bias analysis is available in Multimedia Appendix 1), were examined using the Fisher exact test for dichotomous variables. The Mantel-Haenszel chi-square test was used for ordered categorical variables, chi-square test was used for nonordered categorical variables, and Fisher nonparametric permutation test was used for continuous variables. For categorical variables, number (%) is presented. For continuous variables, mean (SD)/median (min-max)/numbers of individuals is presented.

For comparisons between groups and to assess the effect of the intervention on primary and secondary outcome measures, the Fisher nonparametric permutation test was used for continuous variables. Analysis of covariance (ANCOVA) models were calculated to adjust for differences in baseline GAF scores, yielding parameter estimates of between-group differences with adjusted means with 95% CI and adjusted P values. The same statistical procedure was used to adjust for GAF scores when analyzing the first item of MANSA (overall subjective QoL), which was based on the variable mean score. Primary and secondary outcome measures were predefined before the start of the study and were therefore not adjusted for multiple testing. For comparison within groups on primary and secondary outcome measures, the Wilcoxon signed rank test was used.

Missing data amounted to at most n=1 for all measures at 8 weeks and n=1 for a single scale (HADS) at 6 months. Mean substitution was used to account for missing values at posttreatment and follow-up assessments.

Ethics Approval

The Regional Ethical Review Board at the University of Gothenburg, Sweden, approved the study (Dnr: 013-08; T364-10; T645-11). Written informed consent was obtained from all participants during the intake interviews. All participants were deemed to have a level of maturity required to provide consent for themselves.

Results

Baseline Characteristics

Baseline characteristics presented in Table 1 show that the intervention group and the TAU group were equal in most baseline characteristics. However, the intervention group had a lower mean GAF score than the comparison group. This was also reflected in employment status. Table 2 shows that the intervention and TAU groups were also equal for all outcome measures at baseline (Multimedia Appendix 2).

Table 2. Adjusted primary and secondary variables at baseline, after 8 weeks, and 6 months by treatment group.

Variable	Intervention (IBSC[a]; n=24)		Treatment as usual (n=20)		*P* value between groups: adjusted[b]	Mean difference between groups (95% CI); effect size
	Adjusted mean[b] (95% CI)	*P* value within group	Adjusted mean[b] (95% CI)	*P* value within group		
MANSA[c]						
MANSA total score baseline	53.7 (49.8 to 57.6)	N/A[d]	52.8 (48.5 to 57.1)	N/A	.77	N/A
Change baseline to 8 weeks	0.294[e] (−2.86 to 3.45)	.58	−1.99 (−5.39 to 1.41)	.06	.35	2.28 (−2.57 to 7.13); 0.397
Change baseline to 6 months	0.247 (−3.19 to 3.68)	.81	−1.84 (−5.63 to 1.96)	.21	.44	2.08 (−3.27 to 7.44); 0.281
Subjective QoL[f] (MANSA)						
Subjective QoL baseline	4.55 (4.07 to 5.04)	N/A	4.58 (4.05 to 5.12)	N/A	.94	N/A
Change QoL baseline to 8 weeks	0.163[e] (−0.358 to 0.684)	.53	−0.137 (−0.700 to 0.425)	.48	.45	−0.300 (−0.501 to 1.10); 0.279
Change subjective QoL baseline to 6 months	0.268 (−0.301 to 0.821)	.50	−0.062 (−0.682 to 0.558)	.89	.46	0.321 (−0.554 to 1.20); 0.160
ROSENBERG[g]						
Rosenberg total score baseline	17.3 (15.2 to 19.4)	N/A	17.1 (14.8 to 19.4)	N/A	.90	N/A
Change total score baseline to 8 weeks	1.57[e] (−0.06 to 3.20)	.06	0.043 (−1.72 to 1.81)	.79	.23	1.53 (−0.98; 4.04); 0.332
Change total score baseline to 6 months	1.29 (−0.36 to 2.93)	.04	0.356 (−1.46 to 2.17)	.86	.47	0.930 (−1.64 to 3.49); 0.388
HADS Anx[h]						
HADS Anx baseline	8.02 (6.34 to 9.69)	N/A	7.93 (6.08 to 9.78)	N/A	.95	N/A
Change HADS Anx baseline to 8 weeks	−0.432[e] (−1.68 to 0.813)	.44	1.80 (0.45 to 3.14)	.01	.02	−2.23 (−4.15 to −0.31); 0.796
Change HADS Anx baseline to 6 months	−1.52[e] (−2.85 to –0.20)	.003	1.55 (0.12 to 2.99)	.02	.004	−3.08 (−5.14 to −1.02); 1.24
HADS Depr[i]						
HADS Depr baseline	3.93 (2.57 to 5.29)	N/A	4.48 (2.98 to 5.99)	N/A	.60	N/A
Change HADS Depr baseline to 8 weeks	−0.502[e] (−1.77 to 0.760)	.41	1.58 (0.22 to 2.94)	.06	.04	−2.08 (−4.02; −0.14); 0.675
Change HADS Depr baseline to 6 months	−0.08[e] (−1.38 to −1.22)	.80	1.14 (−0.26 to 2.55)	.09	.23	−1.22 (−3.24; 0.79); 0.491
MADRS-S[j]						
MADRS-S total score baseline	12.0 (9.2 to 14.8)	N/A	12.2 (9.1 to 15.3)	N/A	.94	N/A
Change total score baseline to 8 weeks	−0.103[e] (−2.66 to 2.45)	.96	1.76 (−1.00 to 4.51)	.23	.34	−1.86 (−5.79 to 2.07); 0.272
Change total score baseline to 6 months	−0.243 (−3.18 to 2.70)	.62	1.04 (−2.20 to 4.29)	.33	.57	−1.28 (−5.86 to 3.30); 0.374
SOC[k]						

Variable	Intervention (IBSC[a]; n=24)		Treatment as usual (n=20)		P value between groups: adjusted[b]	Mean difference between groups (95% CI); effect size
	Adjusted mean[b] (95% CI)	P value within group	Adjusted mean[b] (95% CI)	P value within group		
SOC total baseline	125 (116.7 to 133.2)	N/A	121 (112.1 to 130.4)	N/A	.56	N/A
Change SOC total baseline to 8 weeks	2.02[e] (−4.32 to 8.35)	.58	−2.77 (−9.60 to 4.07)	.14	.33	4.78 (−4.96 to 14.5); 0.226
Change SOC total baseline to 6 months	5.81 (−0.62 to 12.2)	.10	−3.39 (−10.5 to −3.71)	.19	.07	9.19 (−0.83 to 19.2); 0.712

[a] IBSC: internet-based support and coaching intervention.

[b] Adjusting for Global Assessment of Functioning interval using analysis of covariance.

[c] MANSA: Manchester Short Assessment of Quality of Life.

[d] N/A: not applicable.

[e] Based on 23 individuals.

[f] QoL: quality of life.

[g] ROSENBERG: Rosenberg Self-Esteem Scale.

[h] HADS Anx: Hospital Anxiety and Depression Scale.

[i] HADS Depr: Hospital Anxiety and Depression Scale.

[j] MADRS-S: Montgomery-Åsberg Depression Rating Scale-Self-reported.

[k] SOC: sense of coherence.

Dropouts

There were a total of 6 dropouts in the intervention group who had completed at least 1 chat session but did not finalize the full 8 weeks of participation, leaving 80% (24/30) of individuals who completed the IBSC. The reasons stated by these participants for dropping out were related to stress at school or at work and/or difficulties in prioritizing and remembering planned sessions. For one individual, reasons for dropping out are not known. No demographic or outcome variables were significantly associated with the probability of dropout (Multimedia Appendix 1 shows the results of the attrition bias analysis).

Effect of Intervention

Table 2 illustrates the adjusted primary and secondary outcome measures at baseline, 8 weeks, and 6 months for the 2 groups (Multimedia Appendix 2 provides further details including unadjusted values and standard deviations).

Primary Outcome

No statistically significant between-group differences were observed in quality of life (MANSA) from pre- to postintervention (8 weeks; adjusted *P*=.35; adjusted mean difference 2.28; 95% CI −2.57 to 7.13) or at the 6-month follow-up (adjusted *P*=.44; adjusted mean difference 2.08; 95% CI −3.27 to 7.44), and neither were there any significant within-group changes in this measure.

Secondary Outcomes

Analyses with ANCOVA adjusted for differing baseline GAF scores revealed a statistically significant between-group difference from baseline to 8 weeks (postintervention) on the HADS (adjusted *P*=.02; adjusted mean difference −2.23; 95% CI −4.15 to −0.31) measuring anxiety symptoms. This difference was explained by a small nonsignificant within-group decrease in anxiety in the intervention group (adjusted mean change score −0.432; 95% CI −1.68 to 0.813) and a significant increase in anxiety in the comparison group (*P*=.01; adjusted mean change score 1.80; 95% CI 0.45 to 3.14). The between-group effect size was large (Cohen *d*=0.80). When comparing the results for anxiety at baseline with those obtained at the 6-month follow-up, there was similarly a statistically significant between-group effect (adjusted *P*=.004; adjusted mean difference −3.08; 95% CI −5.14 to −1.02) with a large between-group effect size (Cohen *d*=1.24). This was explained by a significant decrease in the intervention group (*P*=.003; adjusted mean –1.52; 95% CI −2.85 to −0.20). There was a corresponding significant increase in the comparison group (*P*=.02; adjusted mean 1.55; 95% CI 0.12 to 2.99). For depressive symptoms (according to HADS), there was a significant between-group effect at 8 weeks postintervention (adjusted *P*=.04; adjusted mean difference −2.08; 95% CI −4.02 to −0.14) explained by a nonsignificant decline in depressive symptoms in the intervention group (adjusted mean −0.502; 95% CI −1.77 to 0.760) and a nonsignificant increase in the comparison group (adjusted mean 1.58; 95% CI 0.22 to 2.94) producing a medium-sized between-group effect size (Cohen *d*=0.68).

Finally, there was a significant increase in self-esteem (RSES) for the intervention group at 6 months (*P*=.04; mean change score 1.54, SD 3.59). MADRS-S, measuring depressive symptoms, and MANSA item 1 (subjective QoL) did not improve significantly over time.

Discussion

Principal Findings

This study aimed to investigate whether IBSC, including 2 chat sessions in a week and 2 complementary clinic visits over the course of 8 weeks, could be a feasible support option for young people with ADHD and/or ASD. We found that self-esteem increased and symptoms of anxiety decreased in the intervention group at follow-up. Anxiety and depression had improved postintervention compared with TAU, but these findings were partially due to a deterioration in the TAU group. The primary outcome variable, QoL, did not improve over time compared with the TAU group.

The analysis of intervention effects showed that the intervention group experienced significantly increased self-esteem at follow-up. Overall, it seems that living with ASD and ADHD can lead to a number of adverse consequences that, especially when left unmanaged, increase the risk of developing low self-esteem [48-51]. It also appears that self-esteem in individuals with ASD may be less susceptible to, or take longer to change, following treatment interventions with this objective [52,53]. In individuals with ADHD, some results have suggested that psychoeducation might actually worsen self-esteem initially [54,55], whereas a few later studies have been able to observe preserved or even increased self-esteem following such interventions [56,57]. The above results might be seen in the light of individuals with NDDs having trouble producing appropriate coping strategies [52] and the necessity of including elements of strategy building in approaches with the intention of increasing self-awareness [56,57]. The combination of increased self-knowledge and the acquisition of strategies to better handle the struggles of daily living was the main objective of IBSC. In view of the previously outlined research, it seems reasonable that an increase in self-esteem might occur from a longer perspective rather than directly after the intervention. This is supported by results from the validation study where effects on self-esteem were first seen at the 6-month follow-up [35] as well as from the qualitative study of this intervention, in which participants stated improved self-confidence as an experienced long-term effect [31].

Another observed effect in the intervention group was decreased symptoms of anxiety at follow-up. Individuals with NDDs are at great risk of developing anxiety disorders and have exceedingly elevated rates of lifetime psychiatric comorbidity, which adds to their challenges and may overshadow the NDD [58-60]. In the qualitative study of IBSC, participants reported at least short-term calming effects on emotions, in part related to having someone to turn to with thoughts and questions [31]. There was a tendency for reduced anxiety directly following the intervention, and it is possible that the increase in self-esteem observed between the intervention and at the 6-month follow-up for the intervention group was paralleled by a decrease in anxiety at this measuring point. In this study, the comparison group experienced a significant increase in anxiety both at the 8-week and 6-month measuring points. It is important to address the common issue of comorbid anxiety in this population, and it seems that IBSC might have had a positive effect on this measure over time.

We did not, as hypothesized, observe an effect on our primary outcome measure, quality of life (MANSA), nor did we see an effect on sense of coherence (which relates to subjective QoL) [12,61]. QoL was chosen as the primary outcome measure in keeping with the validation study. No significant improvement in the total score of this instrument was seen in the validation study, but there was a significant increase in subjective QoL as well as an increase in sense of coherence at the 6-month follow-up [35]. There are very few studies of SOC or MANSA in relation to ASD or ADHD, making it difficult to draw any safe conclusions about the results. Another study examined ASD traits and how they relate to daily functioning and specific domains of QoL (as measured by MANSA) [62]. The results of this study showed that even though total ASD symptom severity may correlate (negatively) with overall QoL, there is a complexity in that specific ASD traits and symptoms (eg, degree of insistence on sameness) may have different impacts on separate QoL domains. This suggests that it might be difficult to predict if an intervention will have an impact on total QoL scores, as it may also be funneled by secondary variables. In this study, we did not have access to measures of severity regarding specific ASD or ADHD traits; therefore, we cannot rule out that the current sample may have differed from that of the validation study in a way that might have affected the outcome of these measures. In future studies, it might be wise to consider an alternative primary outcome measure. There are several options that could be considered, one being trying a different QoL self-rating scale, for example, the Short Form Health Survey 36 with specific subscales for mental health [63]. Another option could be to use the Clinical Global Impression Scale, which measures the degree of symptom severity and changes over time, as assessed by a blind assessor at baseline, postintervention, and after 6 months [64].

With regard to SOC, there is only one other intervention study, apart from the validation study, which has made use of this instrument with individuals with NDDs [53]. In this study of a randomized controlled trial comparing group CBT with group recreational activity for adults with ASD, there were no observed effects on the SOC directly after the intervention, even with an increased quality of life. When comparing this study to the validation study, total scores on SOC and MANSA at baseline were more or less equivalent. The difference between our sample and that of the validation study is, however, the age range, where the current sample is older, including individuals up to 32 years. It has been proposed that SOC stabilizes and thus fluctuates less after early adulthood [65]. This might be one explanation for the lack of, or unpredictable, effects on this measure compared with the validation study. Moreover, several of the questions in the SOC partially tap into areas that are associated with core difficulties for individuals with ASD, such as the ability to understand social contexts, perceived predictability, and central coherence. This may also have contributed to a certain degree of resilience in this measure.

In a qualitative study of this intervention [31], participants mentioned several features that were experienced as positive about receiving IBSC: an appreciation of the text-based format, always having someone to turn to when experiencing obstacles, and that coaches had sufficient training and experience in the field of NDDs. Participants also described the intervention as easy-going, practical, and supportive in nature [31]. As mentioned in the introduction, it has been suggested that there is a need for support targeting various basic needs for individuals with NDDs [25]. Such support is recommended in the policy programs of several countries although there is a recognized lack of research in this area [25]. In a multicenter, randomized clinical trial by Philipsen et al [66], individual clinical management (CM) was shown to be as effective as cognitive behavioral group psychotherapy on primary outcome measures (ADHD symptoms). CM consisted of nonspecific individual counseling sessions held by a physician with competence within the field of NDD, addressing themes and issues as requested by the patient [66]. This is not unlike the individualized coaching format of the IBSC. In a qualitative study by Giarelli et al [67], perceived bridges and barriers to successful transitioning for adolescents and young adults with ASD were examined. Participants expressed that a supportive mentor was important and helpful for them in a transitioning process. Overall, it seems that individualized support focusing on aspects of daily living offered by someone with sufficient knowledge might be equally important to narrower skills-based approaches for individuals with NDDs. Offering this kind of support through an internet-based contact ameliorates problems related to sensory difficulties, deciphering nonverbal cues and offering flexibility that might be essential in making support accessible for this group of patients.

All these results suggest that the intervention could be a feasible option for young individuals with NDDs. However, 20% (6/30) of the participants did not complete the IBSC and stated that reasons for this were often stress (eg, due to obligations at work or in studies) or trouble prioritizing sessions. It is possible that the intervention could need further personalization, for example, in relation to the contact frequency, to be able to accommodate differential needs—a fact also indicated by the qualitative study [31]. However, it cannot be ruled out that ADHD alone or in combination with autism could be the reason for the majority of the dropouts of the intervention group. Several previous intervention studies on individuals with ADHD have reported high dropout rates [68].

Strengths and Limitations

This study has limitations that need to be considered. First, groups were not randomized due to expected recruitment problems (ie, individuals with ASD have trouble tolerating uncertainty about the arm they will be allocated to, which was a problem in the undertaking of the validation study). It was also deemed unethical to not provide the control group with adequate support during the intervention period. However, this was a controlled study, and the groups were matched as far as possible by age, sex, and diagnosis. Problems with recruitment at the end of the process meant that some exceptions were made but the analysis of demographics and sample characteristics at baseline showed that there were no differences between groups on these parameters. A retrospective GAF assessment revealed that the intervention group had a significantly lower mean GAF score than the comparison group. In line with this, they were less likely to be employed. To assess the feasibility of the intervention, we tried to recruit a naturalistic sample. This meant that recruitment and intervention took place in routine care settings. Inclusion criteria were deliberately generous including individuals with both ASD and ADHD and only excluding those with serious psychiatric comorbidity (if in need of other treatment). It cannot be entirely ruled out that the intervention group was recruited based on perceived need for support to a larger extent than the comparison group and could be a possible explanation for the difference in GAF scores. The difference was statistically adjusted for the analysis.

As the recruited sample was heterogeneous with regard to diagnosis and age, it might limit the extent to which results are applicable to subsamples such as individuals with exclusively ADHD, ASD, or a certain age range. We suspect that some aspects of the format, such as it being text based and delivered in one’s own home, might have appealed more to individuals with ASD. However, the heterogeneity of the sample is also considered a strength of the study, as it is largely representative of a clinical reality. In recent years, this particular topic has been raised within the context of evidence-based medicine [69]. For example, it has been questioned to what extent results from randomized controlled efficacy studies with narrow inclusion criteria, and often high-functioning participants treated within university settings, are generalizable to clinical contexts [69,70]. We seldom see clear-cut case presentations in clinical practice due to the vast overlap among NDDs and their high frequency of comorbidity [9]. Furthermore, it is important to address the needs of individuals with complex case presentations. They may be in particular need of attendance, but there is a deficiency in research for this group [11,71]. It has also been proposed that adding qualitative measures rather than exclusively relying on quantitative self-report questionnaires could increase the clinical relevance of conducted research and offer an important patient perspective [69].

The results of this study were compiled after some time had elapsed. The reason was partly related to practical circumstances and partly due to the fact that a qualitative study pertaining to the same project had been undertaken, analyzed, and published before the quantitative results [31]. With regard to advances in internet-delivered treatment during this intervening period, there has been a surge of studies in this area of research, including some papers covering synchronous text-based interventions [27,72]. However, these have mainly pertained to interventions such as guided or unguided internet-delivered cognitive behavioral therapy for mental health disorders and to a lesser amount interventions specifically targeting individuals with NDDs. Relevant internet-delivered intervention studies pertaining to this specific group are mentioned in the introduction section. None of the interventions were chat based.

Conclusions

In this IBSC study, an increase in self-esteem and a decrease in anxiety levels were observed 6 months after the intervention. IBSC seems to have significance and is a feasible option for young people with NDDs. There is a need to develop support options that address the obstacles experienced by individuals with NDDs in receiving health care. Future studies should determine in detail for whom and to what degree this kind of support will be regarded as helpful.

Acknowledgments

The authors would like to thank all of the participants and staff at the enrolled clinics for their help and participation. The coaches Kristina Tjernberg, Trine Fohlström, Kicki Lamberg Kastor, Carl Lösnitz, Cristina Dahl, and project coordinator Lena Johannesson are gratefully acknowledged. In addition, the authors thank Karin Martinsson, head of the Neuropsychiatric Treatment Unit, Sodra Alvsborgs Sjukhus, Vastra Götaland Region, and her staff as well as the administrative staff at the habilitation center in Skövde for their assistance. They are grateful to Nils-Gunnar Pehrsson and Anders Pehrsson for assistance in the statistical analyses. The research was made possible by grants from the following research funds: The Health and Medical Care Committee of the Regional Executive Board, Region Västra Götaland; The Health and Medical Care Committee of the Regional Executive Board, Region Halland; The Vårdal Institute; the Swedish Institute for Health Sciences. The authors also acknowledge the financial support of the Fredrik and Ingrid Thurings Foundation, the Jebner foundation, and the Anna-Lisa and Bror Björnsson Foundation. The funding sources had no role in the study design, collection, analysis, and interpretation of data, writing of the report, or the decision to submit the paper for publication.

Authors' Contributions

EW initiated the idea for the study, developed the IBSC model, and secured funding and ethical approval. EW, HS, and IB administered the study. HS and IB recruited participants and collected the data. HS analyzed the data and interpreted the results with significant input from EW, GA, BA, and IB, HS, EW, GA, BA, and IB were involved in writing the paper and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Attrition bias analysis.

[DOCX File , 22 KB - jmir_v22i12e19658_app1.docx]

Multimedia Appendix 2

Primary and secondary outcome measures at baseline, 8 weeks, and 6 months for both study arms (internet-based support and coaching and treatment as usual).

[DOCX File , 24 KB - jmir_v22i12e19658_app2.docx]

References

1. Lundström S, Reichenberg A, Anckarsäter H, Lichtenstein P, Gillberg C. Autism phenotype versus registered diagnosis in Swedish children: prevalence trends over 10 years in general population samples. Br Med J 2015 Apr 28;350:h1961 [FREE Full text] [doi: 10.1136/bmj.h1961] [Medline: 25922345]
2. Atladottir HO, Gyllenberg D, Langridge A, Sandin S, Hansen SN, Leonard H, et al. The increasing prevalence of reported diagnoses of childhood psychiatric disorders: a descriptive multinational comparison. Eur Child Adolesc Psychiatry 2015 Feb;24(2):173-183. [doi: 10.1007/s00787-014-0553-8] [Medline: 24796725]
3. Safer DJ. Is ADHD really increasing in youth? J Atten Disord 2018 Jan;22(2):107-115. [doi: 10.1177/1087054715586571] [Medline: 26084287]
4. Nylander L, Holmqvist M, Gustafson L, Gillberg C. Attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) in adult psychiatry. A 20-year register study. Nord J Psychiatry 2013 Oct;67(5):344-350. [doi: 10.3109/08039488.2012.748824] [Medline: 23234539]
5. Willcutt EG. The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics 2012 Jul;9(3):490-499 [FREE Full text] [doi: 10.1007/s13311-012-0135-8] [Medline: 22976615]
6. Simon V, Czobor P, Bálint S, Mészáros A, Bitter I. Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis. Br J Psychiatry 2009 Mar;194(3):204-211. [doi: 10.1192/bjp.bp.107.048827] [Medline: 19252145]
7. Kim YS, Leventhal BL, Koh Y, Fombonne E, Laska E, Lim E, et al. Prevalence of autism spectrum disorders in a total population sample. Am J Psychiatry 2011 Sep;168(9):904-912. [doi: 10.1176/appi.ajp.2011.10101532] [Medline: 21558103]
8. Blumberg S, Bramlett M, Kogan M, Schieve L, Jones J, Lu M. Changes in prevalence of parent-reported autism spectrum disorder in school-aged U.S. children: 2007 to 2011-2012. Natl Health Stat Report 2013 Mar 20(65):1-11, 1 p following 11 [FREE Full text] [Medline: 24988818]
9. Gillberg C. The ESSENCE in child psychiatry: early symptomatic syndromes eliciting neurodevelopmental clinical examinations. Res Dev Disabil 2010;31(6):1543-1551. [doi: 10.1016/j.ridd.2010.06.002] [Medline: 20634041]

10. Faraone SV, Asherson P, Banaschewski T, Biederman J, Buitelaar JK, Ramos-Quiroga JA, et al. Attention-deficit/hyperactivity disorder. Nat Rev Dis Primers 2015 Aug 6;1:15020. [doi: 10.1038/nrdp.2015.20] [Medline: 27189265]
11. Ahnemark E, Di Schiena M, Fredman AC, Medin E, Söderling JK, Ginsberg Y. Health-related quality of life and burden of illness in adults with newly diagnosed attention-deficit/hyperactivity disorder in Sweden. BMC Psychiatry 2018 Jul 13;18(1):223 [FREE Full text] [doi: 10.1186/s12888-018-1803-y] [Medline: 30005675]
12. Helles A, Gillberg IC, Gillberg C, Billstedt E. Asperger syndrome in males over two decades: quality of life in relation to diagnostic stability and psychiatric comorbidity. Autism 2017 May;21(4):458-469. [doi: 10.1177/1362361316650090] [Medline: 27233289]
13. Hyman SL, Levy SE, Myers SM, Council on Children With Disabilities, Section on Developmental and Behavioral Pediatrics. Identification, evaluation, and management of children with autism spectrum disorder. Pediatrics 2020 Jan;145(1):-. [doi: 10.1542/peds.2019-3447] [Medline: 31843864]
14. Knouse L, Fleming A. Applying cognitive-behavioral therapy for ADHD to emerging adults. Cogn Behav Pract 2016 Aug;23(3):300-315 [FREE Full text] [doi: 10.1016/j.cbpra.2016.03.008]
15. Mattys L, Noens I, Evers K, Baeyens D. 'Hold me tight so I can go it alone': developmental themes for young adults with autism spectrum disorder. Qual Health Res 2018 Jan;28(2):321-333. [doi: 10.1177/1049732317730329] [Medline: 29082803]
16. Cheak-Zamora NC, Farmer JE, Mayfield WA, Clark MJ, Marvin AR, Law JK, et al. Health care transition services for youth with autism spectrum disorders. Rehabil Psychol 2014 Aug;59(3):340-348. [doi: 10.1037/a0036725] [Medline: 25019309]
17. Treuer T, Chan KL, Kim BN, Kunjithapatham G, Wynchank D, Semerci B, et al. Lost in transition: a review of the unmet need of patients with attention deficit/hyperactivity disorder transitioning to adulthood. Asia Pac Psychiatry 2017 Jun;9(2):-. [doi: 10.1111/appy.12254] [Medline: 27552649]
18. De Crescenzo F, Cortese S, Adamo N, Janiri L. Pharmacological and non-pharmacological treatment of adults with ADHD: a meta-review. Evid Based Ment Health 2017 Feb;20(1):4-11. [doi: 10.1136/eb-2016-102415] [Medline: 27993933]
19. Attention deficit hyperactivity disorder: diagnosis and management. The National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/guidance/ng87 [accessed 2020-04-01]
20. Hazell PL, Kohn MR, Dickson R, Walton RJ, Granger RE, van Wyk GW. Core ADHD symptom improvement with atomoxetine versus methylphenidate: a direct comparison meta-analysis. J Atten Disord 2011 Nov;15(8):674-683. [doi: 10.1177/1087054710379737] [Medline: 20837981]
21. Ahmann E, Saviet M, Tuttle L. Interventions for ADHD in children and teens: a focus on ADHD coaching. Pediatr Nurs 2017 May;43(3):121-131 [FREE Full text]
22. Miranda A, Berenguer C, Colomer C, Roselló R. Influence of the symptoms of attention deficit hyperactivity disorder (ADHD) and comorbid disorders on functioning in adulthood. Psicothema 2014;26(4):471-476. [doi: 10.7334/psicothema2014.121] [Medline: 25340893]
23. Autism spectrum disorder in under 19s: support and management. National Institute for Health and Care Excellence. 2013 Aug. URL: https://www.nice.org.uk/guidance/cg170/chapter/1-recommendations [accessed 2020-04-01]
24. Autism spectrum disorder in adults: diagnosis and management. National Institute for Health and Care Excellence. 2012. URL: https://www.nice.org.uk/guidance/cg142/chapter/1-guidance [accessed 2020-04-01]
25. Lorenc T, Rodgers M, Marshall D, Melton H, Rees R, Wright K, et al. Support for adults with autism spectrum disorder without intellectual impairment: systematic review. Autism 2018 Aug;22(6):654-668. [doi: 10.1177/1362361317698939] [Medline: 28683565]
26. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
27. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
28. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
29. Lenhart AS, Anderson M, Duggan M, Perrin A. Teens, Technology and Friendships. Pew Research Center. 2015 Aug. URL: https://www.pewresearch.org/internet/2015/08/06/teens-technology-and-friendships/ [accessed 2020-04-01]
30. Gillespie-Lynch K, Kapp SK, Shane-Simpson C, Smith DS, Hutman T. Intersections between the autism spectrum and the internet: perceived benefits and preferred functions of computer-mediated communication. Intellect Dev Disabil 2014 Dec;52(6):456-469. [doi: 10.1352/1934-9556-52.6.456] [Medline: 25409132]
31. Sehlin H, Ahlström BH, Andersson G, Wentz E. Experiences of an internet-based support and coaching model for adolescents and young adults with ADHD and autism spectrum disorder -a qualitative study. BMC Psychiatry 2018 Jan 18;18(1):15 [FREE Full text] [doi: 10.1186/s12888-018-1599-9] [Medline: 29347983]

32. Bolic Baric V, Hellberg K, Kjellberg A, Hemmingsson H. Internet activities during leisure: a comparison between adolescents with ADHD and adolescents from the general population. J Atten Disord 2018 Oct;22(12):1131-1139. [doi: 10.1177/1087054715613436] [Medline: 26610742]
33. Pettersson R, Söderström S, Edlund-Söderström K, Nilsson KW. Internet-based cognitive behavioral therapy for adults with ADHD in outpatient psychiatric care. J Atten Disord 2017 Apr;21(6):508-521. [doi: 10.1177/1087054714539998] [Medline: 24970720]
34. Backman A, Mellblom A, Norman-Claesson E, Keith-Bodros G, Frostvittra M, Bölte S, et al. Internet-delivered psychoeducation for older adolescents and young adults with autism spectrum disorder (SCOPE): an open feasibility study. Res Autism Spectr Disord 2018 Oct;54:51-64 [FREE Full text] [doi: 10.1016/j.rasd.2018.07.001]
35. Wentz E, Nydén A, Krevers B. Development of an internet-based support and coaching model for adolescents and young adults with ADHD and autism spectrum disorders: a pilot study. Eur Child Adolesc Psychiatry 2012 Nov;21(11):611-622. [doi: 10.1007/s00787-012-0297-2] [Medline: 22736195]
36. First MS, Gibbon M, Williams J. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID). ePROVIDE - Mapi Research Trust. New York: Biometrics Research Department; 1997. URL: https://eprovide.mapi-trust.org/instruments/structured-clinical-interview-for-dsm-iv-axis-i-disorders [accessed 2020-12-15]
37. First M, Spitzer R, Williams J, Benjamin L. Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II). Washington DC: American Psychiatric Press Inc; 1997.
38. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979 Apr;134:382-389. [doi: 10.1192/bjp.134.4.382] [Medline: 444788]
39. Priebe S, Huxley P, Knight S, Evans S. Application and results of the manchester short assessment of quality of life (MANSA). Int J Soc Psychiatry 1999;45(1):7-12. [doi: 10.1177/002076409904500102] [Medline: 10443245]
40. Björkman T, Svensson B. Quality of life in people with severe mental illness. Reliability and validity of the Manchester short assessment of quality of life (MANSA). Nord J Psychiatry 2005;59(4):302-306. [doi: 10.1080/08039480500213733] [Medline: 16195135]
41. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
42. Antonovsky A. Health, stress, and coping. San Francisco: Jossey-Bass Inc; 1979.
43. Antonovsky A. Unraveling the Mystery of Health: How People Manage Stress and Stay Well. San Francisco: Jossey-Bass; 1987.
44. Rosenberg M. Society and adolescent self-image. Princeton, New Jersey: Princeton University Press; 1965.
45. Rosenberg M. Society and the adolescent self-image, revised edition. Middletown, United States: Wesleyan University Press; 1989.
46. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th edn (DSM-IV) ed. Washington, DC: American Psychiatric Press; 1994.
47. Schene A, Van Wijngaarden WB. The involvement evaluation questionnaire. Amsterdam, The Netherlands: Department of Psychiatry, University of Amsterdam; 1992.
48. Tantam D. Psychological disorder in adolescents and adults with asperger syndrome. Autism 2016 Jun 30;4(1):47-62. [doi: 10.1177/1362361300004001004]
49. Jamison TR, Schuttler JO. Examining social competence, self-perception, quality of life, and internalizing and externalizing symptoms in adolescent females with and without autism spectrum disorder: a quantitative design including between-groups and correlational analyses. Mol Autism 2015;6:53 [FREE Full text] [doi: 10.1186/s13229-015-0044-x] [Medline: 26388982]
50. Williamson S, Craig J, Slinger R. Exploring the relationship between measures of self-esteem and psychological adjustment among adolescents with Asperger syndrome. Autism 2008 Jul;12(4):391-402. [doi: 10.1177/1362361308091652] [Medline: 18579646]
51. Cook J, Knight E, Hume I, Qureshi A. The self-esteem of adults diagnosed with attention-deficit/hyperactivity disorder (ADHD): a systematic review of the literature. Atten Defic Hyperact Disord 2014 Dec;6(4):249-268. [doi: 10.1007/s12402-014-0133-2] [Medline: 24668198]
52. Spain D, Blainey SH. Enhancing self-esteem in adults with autism spectrum disorders: a pilot cognitive behaviour therapy (CBT) group intervention. Advances in Autism 2017 Apr 03;3(2):66-75. [doi: 10.1108/aia-06-2016-0018]
53. Hesselmark E, Plenty S, Bejerot S. Group cognitive behavioural therapy and group recreational activity for adults with autism spectrum disorders: a preliminary randomized controlled trial. Autism 2014 Aug;18(6):672-683 [FREE Full text] [doi: 10.1177/1362361313493681] [Medline: 24089423]
54. Wiggins D, Singh K, Getz H, Hutchins D. Effects of brief group intervention for adults with attention deficit/hyperactivity disorder. J Ment Health Couns 1999;21(1):82-92 [FREE Full text]
55. Ratey JJ, Greenberg MS, Bemporad JR, Lindem KJ. Unrecognized attention-deficit hyperactivity disorder in adults presenting for outpatient psychotherapy. J Child Adolesc Psychopharmacol 1992;2(4):267-275. [doi: 10.1089/cap.1992.2.267] [Medline: 19630608]

56. Hirvikoski T, Lindström T, Carlsson J, Waaler E, Jokinen J, Bölte S. Psychoeducational groups for adults with ADHD and their significant others (PEGASUS): a pragmatic multicenter and randomized controlled trial. Eur Psychiatry 2017 Jul;44:141-152 [FREE Full text] [doi: 10.1016/j.eurpsy.2017.04.005] [Medline: 28641216]
57. Hirvikoski T, Waaler E, Lindström T, Bölte S, Jokinen J. Cognitive behavior therapy-based psychoeducational groups for adults with ADHD and their significant others (PEGASUS): an open clinical feasibility trial. Atten Defic Hyperact Disord 2015 Mar;7(1):89-99 [FREE Full text] [doi: 10.1007/s12402-014-0141-2] [Medline: 24863143]
58. McGough JJ, Smalley SL, McCracken JT, Yang M, del Homme M, Lynn DE, et al. Psychiatric comorbidity in adult attention deficit hyperactivity disorder: findings from multiplex families. Am J Psychiatry 2005 Sep;162(9):1621-1627. [doi: 10.1176/appi.ajp.162.9.1621] [Medline: 16135620]
59. Rosen TE, Mazefsky CA, Vasa RA, Lerner MD. Co-occurring psychiatric conditions in autism spectrum disorder. Int Rev Psychiatry 2018 Feb;30(1):40-61. [doi: 10.1080/09540261.2018.1450229] [Medline: 29683351]
60. Davidsson M, Hult N, Gillberg C, Särneö C, Gillberg C, Billstedt E. Anxiety and depression in adolescents with ADHD and autism spectrum disorders; correlation between parent- and self-reports and with attention and adaptive functioning. Nord J Psychiatry 2017 Nov;71(8):614-620. [doi: 10.1080/08039488.2017.1367840] [Medline: 28836480]
61. Eriksson M, Lindström B. Antonovsky's sense of coherence scale and its relation with quality of life: a systematic review. J Epidemiol Community Health 2007 Nov;61(11):938-944 [FREE Full text] [doi: 10.1136/jech.2006.056028] [Medline: 17933950]
62. Deserno MK, Borsboom D, Begeer S, Geurts HM. Relating ASD symptoms to well-being: moving across different construct levels. Psychol Med 2018 May;48(7):1179-1189. [doi: 10.1017/S0033291717002616] [Medline: 28889821]
63. McHorney CA, Ware JE, Lu JF, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994 Jan;32(1):40-66. [doi: 10.1097/00005650-199401000-00004] [Medline: 8277801]
64. Guy W, National Institute of Mental Health (U.S.). Psychopharmacology Research Branch, Early Clinical Drug Evaluation Program. ECDEU assessment manual for psychopharmacology. Rockville, Maryland: U.S. Dept. of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs, 1976; 1976.
65. Eriksson M, Lindström B. Validity of Antonovsky's sense of coherence scale: a systematic review. J Epidemiol Community Health 2005 Jun;59(6):460-466 [FREE Full text] [doi: 10.1136/jech.2003.018085] [Medline: 15911640]
66. Philipsen A, Jans T, Graf E, Matthies S, Borel P, Colla M, Comparison of MethylphenidatePsychotherapy in Adult ADHD Study (COMPAS) Consortium. Effects of group psychotherapy, individual counseling, methylphenidate, and placebo in the treatment of adult attention-deficit/hyperactivity disorder: a randomized clinical trial. JAMA Psychiatry 2015 Dec;72(12):1199-1210. [doi: 10.1001/jamapsychiatry.2015.2146] [Medline: 26536057]
67. Giarelli E, Ruttenberg J, Segal A. Bridges and barriers to successful transitioning as perceived by adolescents and young adults with asperger syndrome. J Pediatr Nurs 2013;28(6):563-574. [doi: 10.1016/j.pedn.2012.12.010] [Medline: 23541737]
68. Soendergaard HM, Thomsen PH, Pedersen P, Pedersen E, Poulsen AE, Nielsen JM, et al. Treatment dropout and missed appointments among adults with attention-deficit/hyperactivity disorder: associations with patient- and disorder-related factors. J Clin Psychiatry 2016 Feb;77(2):232-239. [doi: 10.4088/JCP.14m09270] [Medline: 26761266]
69. Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? Br Med J 2014 Jun 13;348:g3725 [FREE Full text] [doi: 10.1136/bmj.g3725] [Medline: 24927763]
70. Walter D, Dachs L, Faber M, Goletz H, Goertz-Dorten A, Hautmann C, et al. Effectiveness of outpatient cognitive-behavioral therapy for adolescents under routine care conditions on behavioral and emotional problems rated by parents and patients: an observational study. Eur Child Adolesc Psychiatry 2018 Jan;27(1):65-77. [doi: 10.1007/s00787-017-1021-z] [Medline: 28685400]
71. Katzman MA, Bilkey TS, Chokka PR, Fallu A, Klassen LJ. Adult ADHD and comorbid disorders: clinical implications of a dimensional approach. BMC Psychiatry 2017 Aug 22;17(1):302 [FREE Full text] [doi: 10.1186/s12888-017-1463-3] [Medline: 28830387]
72. Andersson G. Internet interventions: past, present and future. Internet Interv 2018 Jun;12:181-188 [FREE Full text] [doi: 10.1016/j.invent.2018.03.008] [Medline: 30135782]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ANCOVA: analysis of covariance
ASD: autism spectrum disorder
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
CM: clinical management
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
GAF: Global Assessment of Functioning

HADS: Hospital Anxiety and Depression Scale
IBSC: internet-based support and coaching
MADRS-S: Montgomery-Åsberg Depression Rating Scale-Self-reported
MANSA: Manchester Short Assessment for Quality of Life
NDD: neurodevelopmental disorder
QoL: quality of Life
RSES: Rosenberg Self-Esteem Scale
SCID I: Structured Clinical Interview for DSM Axis I Disorder
SCID II: Structured Clinical Interview for DSM-IV Axis II Disorder
SOC: Sense of Coherence
TAU: treatment as usual

Edited by G Eysenbach; submitted 27.04.20; peer-reviewed by T Ewais, M Mulvenna; comments to author 18.07.20; revised version received 28.07.20; accepted 22.09.20; published 31.12.20.

Please cite as:
Sehlin H, Hedman Ahlström B, Bertilsson I, Andersson G, Wentz E
Internet-Based Support and Coaching With Complementary Clinic Visits for Young People With Attention-Deficit/Hyperactivity Disorder and Autism: Controlled Feasibility Study
J Med Internet Res 2020;22(12):e19658
URL: http://www.jmir.org/2020/12/e19658/
doi:10.2196/19658
PMID:33382381

Original Paper

Awareness Development and Usage of Mobile Health Technology Among Individuals With Hypertension in a Rural Community of Bangladesh: Randomized Controlled Trial

Yasmin Jahan[1], MPH, MBBS; Md Moshiur Rahman[1], PhD, MPH, MBBS; Abu S G Faruque[2], MPH, MBBS; Mohammod Jobayer Chisti[2], PhD, MMed, MBBS; Kana Kazawa[1], PhD, RN; Ryota Matsuyama[1], PhD, DVM; Michiko Moriyama[1], PhD, RN

[1]Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan
[2]International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh

Corresponding Author:
Yasmin Jahan, MPH, MBBS
Graduate School of Biomedical and Health Sciences
Hiroshima University
1-2-3 Minami-ku, Kasumi-cho, Hiroshima-shi
Hiroshima, 7340046
Japan
Phone: 81 0822575365
Email: dr.yasminjahan@gmail.com

Abstract

Background: Hypertension (HTN) is a major modifiable risk factor and the leading cause of premature deaths globally. The lack of awareness and knowledge have been identified as risk factors in low- and middle-income countries including Bangladesh. Recently, the use of mobile phone SMS text messaging is found to have an important positive impact on HTN management.

Objective: The study aimed to develop awareness and knowledge in order to enhance lifestyle behavior changes among individuals with HTN in a rural community of Bangladesh by using health education and mobile health (mHealth) technology (SMS text messaging).

Methods: A prospective randomized 5-month intervention, open-label (1:1), parallel-group trial was implemented among the individuals with HTN aged 35 years or older. Both men and women were included. Between August 2018 and July 2019, we enrolled 420 participants, selected from a tertiary level health facility and through door-to-door visits by community health workers. After block randomization, they were assigned to either the intervention group (received SMS text messaging and health education; n=209) or the control group (received only health education; n=211). The primary outcome was the evaluation of self-reported behavior changes (salt intake, fruits and vegetables intake, physical activity, and blood pressure [BP], and body weight monitoring behaviors). The secondary outcomes were measurements of actual salt intake and dietary salt excretion, blood glucose level, BP values, and quality of life (QOL).

Results: During the study period, a total of 8 participants were dropped, and the completion rate was 98.0% (412/420). The adherence rates were significantly higher (9%) among the control group regarding salt intake (*P*=.04) and physical activity behaviors (*P*<.03), and little differences were observed in other behaviors. In primary outcome, the focused behavior, salt intake less than 6 g/day, showed significant chronological improvement in both groups (*P*<.001). The fruits intake behavior steadily improved in both groups (*P*<.001). Participants in both groups had a custom of vegetables intake everyday/week. Physical activity suddenly increased and continued until the study end (*P*<.001 in both groups). Both BP and body weight monitoring status increased from baseline to 1 month but decreased afterward (*P*<.001). In case of secondary outcomes, significant chronological changes were observed in food salt concentration and urinary salinity between the groups (*P*=.01). The mean systolic BP and diastolic BP significantly chronologically decreased in both groups (systolic BP, *P*=.04; diastolic BP, *P*=.02.*P*<.05). All of these supported self-reported behavior changes. For the QOL, both groups showed significant improvement over the study periods (*P*<.001).

Conclusions: Based on these results, we suggest that face-to-face health education requires integration of home health care provision and more relevant and timely interactive SMS text messages to increase the effectiveness of the intervention. Besides, community awareness can be created to encourage "low-salt culture" and educate family members.

Trial Registration: Bangladesh Medical Research Council (BMRC) 06025072017; ClinicalTrials.gov NCT03614104; https://clinicaltrials.gov/ct2/show/NCT03614104 and UMIN-CTR R000033736; https://tinyurl.com/y48yfcoo

International Registered Report Identifier (IRRID): RR2-10.2196/15523

(J Med Internet Res 2020;22(12):e19137) doi:10.2196/19137

KEYWORDS

mobile health; hypertension; behavior changes; awareness development; lifestyle; Bangladesh

Introduction

Noncommunicable diseases (NCDs) are the leading cause of death and disability globally. Each year an estimated 41 million people die from NCDs, accounting for about 70% of total mortality worldwide [1]. Among the NCDs, hypertension (HTN) is a major modifiable risk factor and the leading cause of premature death [2]. The global prevalence of HTN is projected to increase up to 29.2% by 2025 (or >1.5 billion people) [3-5]. Currently, the prevalence of HTN [3] among adults is higher in low- and middle-income countries (31.5% or 1.04 billion people) than in high-income countries (28.5% or 349 million people) [6]. A recent report suggested that 18% or 12 million adults aged 25 years or older suffer from HTN in Bangladesh [7].

The risk factors for developing HTN are more salt intake, sedentary lifestyle, excess weight, unhealthy diet, tobacco consumption, and chronic stress [8-10]. The World Health Organization recommends salt intake of 5 g/day for adults in order to reduce blood pressure (BP) and risk of cardiovascular diseases [11], whereas Bangladeshi people consume more than 17 g salt/day [12].

According to the NCD risk factors survey, the prevalence of self-reported HTN is 12.5% (men 10.9% and women 13.9%) and one-third of the Bangladeshi population never measured their BP due to lack of opportunity and accessibility to the health care system [13]. Despite the increasing prevalence of HTN, a substantial proportion of the population remains undiagnosed due to lack of awareness to seek medical care and high treatment costs [7,11,14]. Controlling BP at a population level may be associated with a decrease in cardiovascular diseases regardless of the increasing burden of chronic diseases such as kidney disease and diabetes [15]. The comorbid conditions related to HTN may cause the greatest morbidity burden if left untreated, especially in low-resource settings [3,4]. Therefore, continuous and long-term BP monitoring including medication adherence is needed [16,17]. So far, different types of nonpharmacological interventions have been developed and tested with aims to improve BP control [18].

In this regard, the 2-way SMS text messaging communication is a commonly used approach that can help develop awareness and improve an individual's self-management about HTN [19]. Nowadays, the mobile phone has become the favored device to communicate and gather information together worldwide [19]. Mobile phone–based interventions could address individual-level factors in health by facilitating timely patient access to relevant health information and support, making patient–provider communication easier, and providing context-specific support that translates into action [19]. Moreover, SMS text messaging is available on all mobile phone service providers that support 1- or 2-way communication [20]. As an example, some trials suggested that SMS text messaging reminders may improve behavior, clinic attendance, attention in care, and self-reported medication adherence [21]. Studies have explored the efficacy of SMS-based interventions for people with HTN to improve treatment compliance and BP control [22,23]. Although the results of trials suggested the potential for small to moderate effect sizes in improving compliance and BP control, the effects were not statistically significant [20]. Furthermore, systematic reviews and meta-analysis on SMS text messaging–based interventions provide researchers useful insight into the most promising applications of SMS text messaging in supporting health care, disease prevention, and case management. Despite the general interest in mobile phone–based interventions related to disease management and behavior change, the allied evidence remains limited [20]. However, while a significant number of studies focused on chronic disease prevention, particularly cardiovascular disease prevention by utilization of SMS text messaging, or on diabetes management [19], there are limited studies on BP outcomes [24].

To fill the gap, in this study, we provided periodic reminders via SMS text messaging along with in-person health education aimed at increasing awareness about BP level and behavior changes to positively influence patients' perceptions of HTN and their adherence to treatment remedies [25]. Participants from the rural community were selected for the study based on 3 criteria: (1) have lack of awareness (understanding) about HTN due to low health literacy, (2) do not have sufficient health infrastructure, and (3) lower adherence rate compared to urban community people [7,26]. Hence, to expand awareness of the target population, 3 innovative techniques (Portable Health Clinic, food and urine salinometer, and SMS text messaging) were introduced in this study [27].

The purpose of the study was to develop awareness and knowledge in order to enhance lifestyle behavior changes among individuals with HTN in a rural community of Bangladesh by using mobile health (mHealth) technology (SMS text messaging) and in-person health education. Thus, this approach could minimize the hypertensive status of the participants.

Methods

Study Design

This was a single-center, prospective randomized (1:1), open-label, parallel-group study conducted in a rural community of Bangladesh involving people with HTN. The intervention period was 5 months, and the total study duration was 12 months. The detailed study design has been previously reported [27]. The study is reported according to the Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth V1.6.1 (Multimedia Appendix 1) [28].

Study Population and Sampling

This study was conducted between August 2018 and July 2019, and we enrolled a total of 420 participants aged between 35 and 71 years. Individuals with HTN were identified from a tertiary level health facility by reviewing the registered clinician's prescriptions and from the hospital logbook. Further, community health workers (CHWs) who received information from the hospital did a door-to-door visit to select participants. The female CHWs were selected from the study community and trained for data collection, including providing in-person health education and SMS text messaging. A purposive sampling method was used for enrolling participants who met the eligibility criteria and were approached to participate voluntarily in the study. Participants (either sex) enrolled in this study who were diagnosed with HTN were aged 35 years or older, had 1-5 years of schooling, resided within a radius of 3 miles from the tertiary hospital, decided to stay in the community for at least five months, had a personal cell phone or access to a shared phone, could exchange views freely, and were willing to participate in the study. Individuals with mental illnesses or serious comorbidities such as malignancy or tuberculosis were excluded from this study.

Sample Size

The sample size calculation was based on behavior changes of the study participants and has been described previously [27]. The sample size was calculated with a 2-tailed 5% significance level, 80% power with a CI of 95% (1–α), and to detect varying differences in the effectiveness of the 2 intervention groups. Adherence rates were assumed to be differing between 10% and 12% with 90% in the study group with a presumption that 6% of the participants would be lost during follow-up. Thus, considering the largest size of the calculated sample sizes, the study finally had a sample size of 210 in each group.

Randomization

A randomization schedule was prepared following a permuted block randomization technique using a block size of 4 based on a computer-generated series of numbers. The serially numbered sealed envelopes were used to allocate the participants to either the intervention (n=209) or the control (n=211) group after written informed consent was obtained.

Study Procedure

The study procedure is described in the following sections and was also described in detail previously [27] (also see Multimedia Appendix 2).

Intervention Group

The intervention group received 5 months' in-person health education along with a health education booklet and SMS text messaging to develop awareness and knowledge, and motivate them for behavior changes, with the content of both educational materials and SMS text messaging being the same.

We developed health education materials, SMS text messaging, and questionnaire according to the Dietary Approaches to Stop Hypertensive (DASH) diet and the National Institutes of Health–and the World Health Organization–recommended guidelines [29]. All materials were translated to Bangla for the participants. The SMS text messaging served as a reminder for behavior changes based on the DASH diet. The contents of the SMS text messages included physical activity, medication adherence, less salt intake (ie, salt intake <6 g daily), eat more fruits and vegetables and less beef and mutton, avoid junk foods, do physical activity for 30 minutes at least five days a week, and take antihypertensive medication regularly [27]. The contents of SMS text messages are shown in Multimedia Appendix 3.

At the time of enrollment, CHWs visited the participant's household in the morning. The CHWs conducted interview and performed physical examination (height, weight, and BP) and blood test (random blood sugar [RBS]) of the study participants at the same time. Furthermore, the CHWs checked food salinity by measuring salt in liquid foods and urinary salinity (spot test [only at baseline] and first-morning urine [overnight urine]). They also checked the urinary protein levels. Afterward, the CHWs provided in-person health education to change participant's behaviors. Participants of the intervention group were followed up every month up to 5 months (two times in the first month and one time for the rest). The SMS text messages were sent 5 times for the first month and one time a week for the remaining 4 months (a total of 21 SMS text messages).

Control Group

The control group received the same health education booklet as the intervention group at the time of enrollment. They were followed up similarly every consecutive month up to 5 months (two times in the first month and one time for the rest). However, they received only in-person health education identical to that of intervention group and not the SMS text messages.

Outcome Measures

The primary outcome was the evaluation of self-reported lifestyle behavior changes (salt intake, fruits intake, vegetables intake, physical activity, BP monitoring, and body weight monitoring). The measurement tool was a Likert-type response scale and evaluated lifestyle behavior changes every month on a scale from 1 to 5, where 1 demonstrates "0 days/week (approximately 13-14 g daily salt intake)" in case of salt intake, or "0 days/week" in case of fruits intake, vegetables intake, physical activity, or "Never/month" (BP and body weight monitoring)" and 5 implies the "Everyday/week (<6 g daily salt intake)," "Everyday/week" (fruits intake, vegetables intake, physical activity), or "8 times/month" (BP and body weight monitoring).

Regarding salt intake, CHWs asked participants to fill their spoons with the salt they usually used and measured approximately how many grams they were taking per day. In case of fruits and vegetables intake, consuming a single piece of fruit and at least one meal with vegetables everyday/week corresponds to the Likert-type response scale score of 5. However, the quantities of fruits and vegetables intake were not measured. With regard to physical activity, participants performing at least 30 minutes of physical activity everyday/week were given a score of 5 on the Likert-type response scale. The detailed content of the behavior change questionnaire is presented in Multimedia Appendix 4. The behavior change was measured by adherence rates of each behavior which was operationally defined as the improvement of behavior (ie, any shift to the higher score on the Likert-type response scale in case of each behavior) at 5 months compared with that at the baseline.

The secondary outcomes were (1) actual salt intake measured by food salinity checker (TANITA electron salinometer SO-313) and dietary salt excretion measured by urine (KME-03, KOUNO ME Institute) [30]. We measured the salt concentration of food using 3 different categories (yellow indicated moderate [0.4%-0.7%], green indicated thick [0.8%-1.1%], and red indicated very thick [1.2%–1.4%] considering 100 mL of food); BP values; RBS level; urinary protein excretion measured by urinalysis strip test for checking comorbidities (ie, kidney diseases and diabetes). The strip analysis usually detects protein concentration in urine. The results were graded as negative, trace (10-20 mg/dL), 1+ (30 mg/dL), 2+ (100 mg/dL), 3+ (300 mg/dL), or 4+ (1000 mg/dL) [29]. The quality of life (QOL) was measured using the standardized EuroQol-5 dimensions-5-level (EQ-5D-5L) questionnaire, which was developed for measuring 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) [31,32] and each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and unable to/extreme problems. A higher score indicated better QOL [33]. In this study, we set the standard value (tariff) for EQ-5D-5L to estimate the impact of health care interventions on QOL (used Japanese region's tariff) [34].

Ethical Consideration

This study was approved by the Bangladesh Medical Research Council (BMRC; Approval No. 06025072017) and registered with the Clinical Trial Registry (NCT03614104). The UMIN Registration Number is R000033736. This study was conducted in accordance with the Declaration of Helsinki and the Ethical Guidelines on Clinical Studies of the Ministry of Health, Labor and Welfare of Japan.

All participants had been explicitly explained about the objectives, importance, risks, and benefits of the research before recruitment. Participation was completely voluntary and written informed consent was obtained from all participants.

Quality Control

The data quality control has been explained in detail in our previous work [27].

Statistical Analysis

In order to examine the effectiveness of the intervention, baseline and end-line surveys were conducted in both study groups. The intention-to-treat analysis was conducted to compare the outcomes of the intervention and control groups. To ensure the comparability of the randomized samples, all baseline indicators at the time of registration were analyzed for both groups. The data were expressed as the mean (SD) or median (range) and cross-tabulation for the categorical variables and as frequencies and percentages for the discrete variables. The baseline differences in categorical variables between the groups were examined using the chi-square (χ^2) test and for continuous variables using the *t* test or Mann–Whitney *U* test, as appropriate.

For the primary outcome, we have tested the difference between adherence rates of the intervention and control groups for each behavior. The adherence rate (proportion) was calculated as follows: dividing the number of individuals who adhered to the intervention (or control) group by the number of participants in the intervention (or control) group in the 5 months (ie, participants who dropped out from the program were not included in the denominator). Statistical differences were tested by the χ^2 test.

In addition, to evaluate the effect of SMS text messaging plus in-person health education compared to only in-person health education on the participant's behavior, we employed 2 types of statistical model: one assumed that there was no effect of the intervention on the time-series change of the proportion of each Likert-type response scale score, and the other assumed that the intervention affected the time-series change of the proportion of each Likert-type response scale score. In other words, we fit a multinomial logit model to observe the proportional change of each Likert-type response scale score (1) with only the time (baseline, 0.5 months, 1 month, 2 months, 3 months, 4 months, and 5 months) as an explanatory variable (time-only model, which means the chronological effect of intervention did not differ from that of control), and (2) with the time and the type of group (intervention/control) as explanatory variables (time–group model, which means the chronological effect of the intervention was different from that of control). The detailed information about the model structure is described in Multimedia Appendix 5. In the time–group model, the control group was used as the reference and the proportional effects of the intervention were estimated. Except for vegetables intake behavior, we fit the models for segmented times designed for each behavior because the changes in each behavior were not constant (Figure S2 and Tables S1 and S2 in Multimedia Appendix 5). The segmentation was implemented by assuming the constant trends (1) from baseline to 0.5 months, and 0.5 months to 5 months in salt intake and physical activity, and (2) from baseline to 0.5 months, 0.5 month to 1 month, 1 month to 2 months, and 2 months to 5 months in BP and body weight monitoring. We evaluated the goodness of fit for both models by the likelihood ratio test with a significance level (α=.05). In that case, the result of the likelihood ratio test was not significant when we selected the time-only model. In case the time–group model was selected as the better fit model, the relative effects of an intervention on the control group were calculated from the estimated coefficients with 95% confidence intervals (CIs). To obtain maximum likelihood estimates of parameters, we used the statistical software R (R Foundation) [35] and ‘multinom’ function implemented in the package ‘nnet’ [36]. For data manipulation and graphical presentation, we used the packages ‘dplyr’ [37] and ‘ggplot2’ [38]. The strength of association was measured by calculating the risk ratios and their 95% CIs. The Friedman test was performed for the chronological comparison of each behavior.

For secondary outcomes, to compare numeric outcome variables between the 2 groups, 2-way repeated-measures ANOVA was performed after verification of data normality. Moreover, the number of participants for each salt concentration measurement was counted at every month, and the Friedman test was performed.

The wealth index was constructed using principal component analysis of the household assets. The households were classified into socioeconomic status quintiles based on the wealth index: quintile (poor, lower middle, middle, upper middle, and rich).

Data were analyzed using the statistical software packages SPSS for Windows version 25.0 (IBM) and Epi Info version 7.0 (CDC). Two-tailed *P*-values were used, and the level of significance of association was considered at <.05.

Data Availability

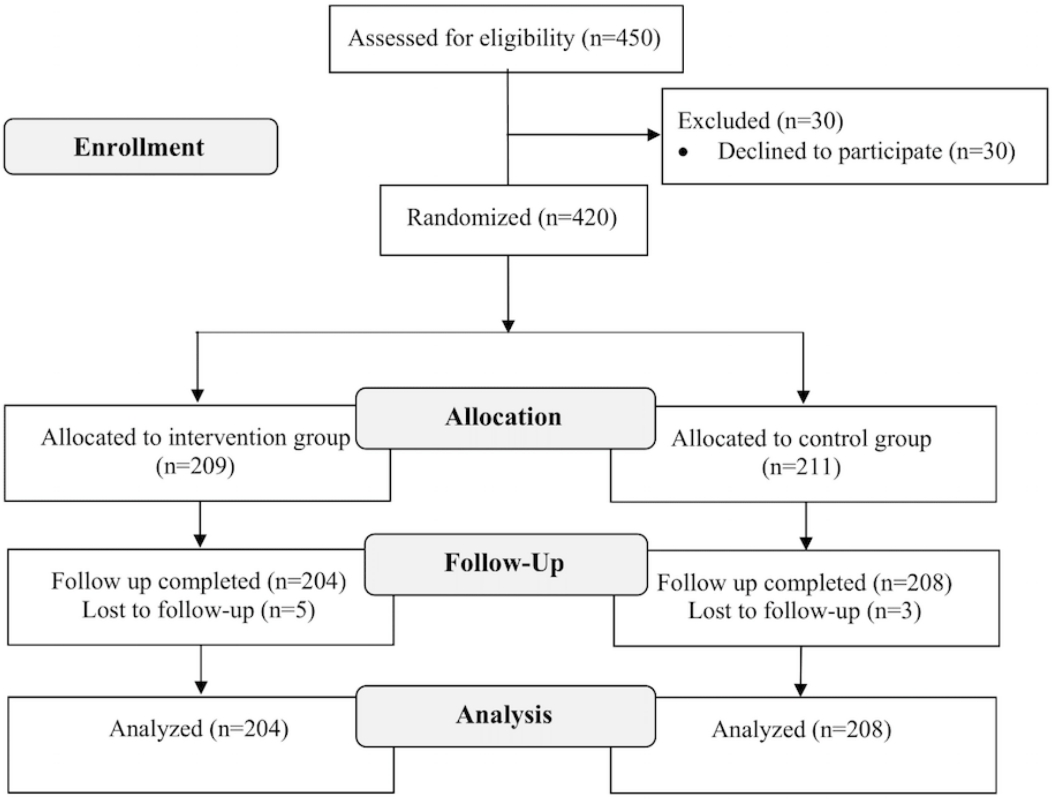
The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Results

Study Population

Of the 450 individuals who met the eligible criteria, 420 consented to participate in the study (consent rate, 93.3%). During the intervention period, 5 from the intervention group and 3 from the control group dropped out due to the refusal to continue. Therefore, 204/209 (97.6%) and 208/211 (98.6%) completed the program in the intervention and control groups, respectively, and their data were analyzed (Figure 1).

Figure 1. Flow chart of the study participants.



Demographic Characteristics

The demographic data and participant profiles of each group are shown in Table 1. In total, the mean (SD) age was 47.1 (8.4) years. As much as 361/420 (86.0%) of the participants were female, and 346/361 (95.8%) were housewives. Most participants had their spouses (378/420 [90.0%]), and 80/420 (19.0%) had an educational background of secondary schooling completed or higher. Moreover, 287/420 (68.3%) of participants were taking their antihypertensive medicine regularly. Regarding tobacco use, 110/420 (26.2%) of the total study participants smoked either smokeless tobacco or cigarettes or both, whereas only 17/420 (4.0%) smoked cigarettes every day. With regard to RBS levels, the average (SD) value was calculated and compared to the intervention group (9.3 [SD 5.5]) and the control group (8.9 [SD 4.8]). As much as 113/204 (55.4%) of participants from the intervention group and 117/208 (56.3%) from the control group had normal or lesser values than those recommend for RBS. Only 4 participants had 1+ (30 mg/dL) and 2 participants had 2+ (100 mg/dL) urinary protein excretion among the total study participants. No statistically significant differences were observed between the 2 groups (Table 1).

Table 1. Demographic characteristics, behavior, and comorbidities of the study participants (N=420).

Variables name	Intervention group (N=209)	Control group (N=211)	*P* value
Age (in years)			.10[a]
Mean (SD)	46.4 (8.3)	47.8 (8.6)	
Range	35-70	35-71	
Gender, n (%)			.26[b]
Female	184 (88.0)	177 (83.9)	
Educational status, n (%)			
Primary complete	97 (46.4)	104 (49.3)	.33[b]
Secondary incomplete	68 (32.5)	71 (33.6)	
Secondary complete or higher	44 (21.1)	36 (17.1)	
Marital status, n (%)			
Married	190 (90.9)	188 (89.1)	.09[b]
Occupation, n (%)			
Housewife	177 (84.7)	169 (80.1)	.46[b]
Wealth index, n (%)			
Poor	46 (22.0)	59 (28.0)	.56[b]
Lower middle	53 (25.4)	53 (25.1)	
Middle	58 (27.8)	46 (21.8)	
Upper middle	44 (21.1)	45 (21.3)	
Rich	8 (3.8)	8 (3.8)	
Antihypertensive drug, n (%)			
Take regularly	151 (72.2)	136 (64.5)	.09[b]
Tobacco use[d] (Everyday), n (%)			
Smoke cigarette	5 (2.4)	12 (5.7)	.08[b]
Smokeless tobacco intake	53 (25.4)	57 (27.0)	.23[b]
Comorbidities			
Random blood sugar (mmol/L)			
Mean (SD)	9.3 (5.5)	8.9 (4.8)	.80[c]
Normal range (4.4-7.8) or below	113 (26.9)	117 (27.9)	.77[b]
Urinary protein (n)			
Trace amount	1	0	.56[b]
1+	3	1	
2+	1	1	

[a] *t* test.

[b] Chi-square test.

[c] Mann–Whitney *U* test.

[d] Duplicate answer.

Primary Outcome (Behavior Change)

Adherence Rate

The adherence rates for the intervention and control groups are presented in Table 2. The adherence rates were significantly higher among the control group participants regarding salt intake behavior (the difference was 9% between the 2 groups; *P*=.04) and physical activity behavior (also 9%; *P*=.03); however, very little differences were observed in other behaviors.

Table 2. Adherence rates in the intervention and the control groups.

Behavior	Adherence rate[a]		Differences	χ^2 test
	Intervention group, %	Control group, %	Intervention – Control	*P* value
Salt intake	66.5	75.8	–9.3	.04
Fruits intake	64.6	66.8	–2.2	.68
Vegetables intake	1.0	0.0	1.0	.25
Physical activity	72.7	82.0	–9.3	.03[b]
Blood pressure monitoring	12.4	13.3	–0.8	.88
Body weight monitoring	1.4	1.4	0.0	>.99

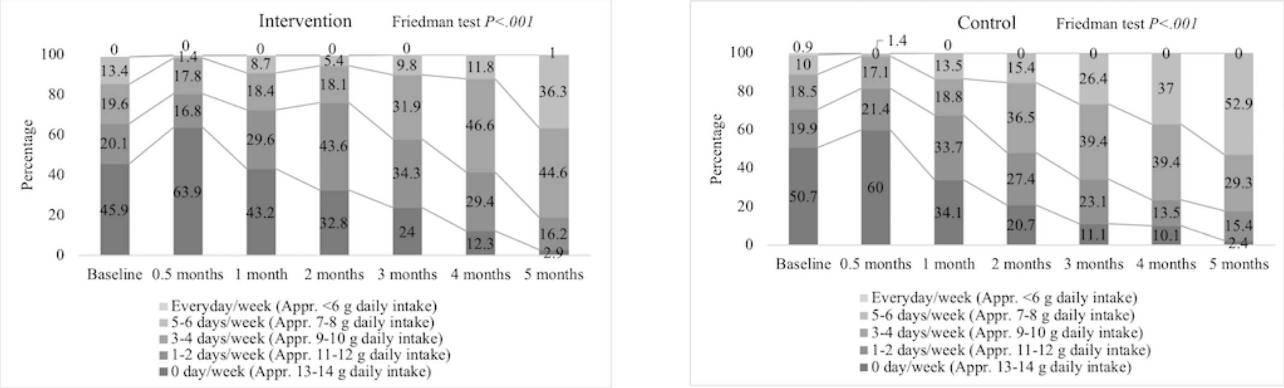
[a] Proportion of behavior improved (number of participants who improved their behavior from baseline to 5 months).

[b] *P* value <.05.

Intervention Effects on Behavior Changes

Regarding behavior changes, the salt intake behavior which we recommended (ie, <6 g daily intake) showed significant improvement in both the intervention and control groups (*P*<.001). The intervention group improved from “0 days/week” (coded as 1) to “5-6 days/week” (coded as 4) from 96/209 (45.9%) at baseline to 6/209 (2.9%) at 5 months and 28/204 (13.4%) at baseline to 74/204 (36.3%) at 5 months, respectively, over the study period. Besides, the control group extensively improved from “0 days/week” to “5-6 days/week” from 107/211 (50.7%) at baseline to 5/208 (2.4%) at 5 months and 21/211 (10.0%) at baseline to 110/208 (52.9%) at 5 months, respectively (Figure 2). The chronological changes in other lifestyle behaviors (fruits intake, vegetables intake, physical activity, BP monitoring, and body weight monitoring) are described in Multimedia Appendix 5 and the respective figures are shown in Multimedia Appendix 6 (Figures S3-S7).

Figure 2. Chronological salt intake behavior changes between the intervention and control group.



The fruits intake steadily improved in both groups, which was statistically significant (*P*<.001 in each group). In both groups, the percentages were increased by around 20% among those participants who ate fruits everyday/week from baseline to the end of the study. Participants in both groups had a custom of vegetable intake everyday/week. There were no changes between the 2 groups, and hence the differences were not significant.

The majority of the participants (140/209 [67.0%] in the intervention and 163/211 [77.3%] in the control group) did not indulge in physical activity at baseline. However, after the intervention, most of them (195/204 [95.6%] in the intervention and 197/208 [94.7%] in the control group) reached the scale “Everyday/week” and continued to take part in some kind of physical activity on a daily basis until the study end (*P*<.001, both groups).

BP and body weight monitoring status per month were increased from baseline to 1 month but decreased afterward. The difference was statistically significant in both groups (*P*<.001). Up to 1 month, CHWs measured BP and body weight two times per month, but afterward one time a month for remaining 4 months (Multimedia Appendix 6).

In accordance with the results of likelihood ratio tests, we selected the time–group model for estimating salt intake (*P*<.001) and fruits intake behavior (*P*<.001; Table 3). Besides, the time–group model was not selected for physical activity, BP, and body weight monitoring as the changes in these behaviors were inconsistent. These results showed that the difference between the effect of intervention on the 2 groups was significant in salt intake and fruits intake than other behaviors.

The intervention consistently contributed to the increase of the Likert-type response scale scores 2 (1-2 days/week) and 5 (Everyday/week) in fruits intake. The relative effect of the intervention ranged from 1.28 (95% CI 1.00-1.61) times increase in score 2 (1-2 days/week) to 2.24 (95% CI 1.68-2.97) times increase in score 5 (Everyday/week). By contrast, in case of salt intake, the intervention resulted in a decrease of the Likert-type response scale scores 3 (approximately 9-10 g daily intake) and 4 (approximately 7-8 g daily intake) in term 2. The relative effects were estimated as 0.63 (95% CI 0.50-0.79) times decrease in score 3 (approximately 9-10 g daily intake) and 0.29 (95% CI 0.22-0.39) times decrease in score 4 (approximately 7-8 g daily intake; Multimedia Appendix 5).

Table 3. Statistical significance in the likelihood ratio test and the selected model for each behavior among the study groups.

Behavior	*P* value	Selected model
Salt intake	<.001	Time–group
Fruits intake	<.001	Time–group
Vegetables intake	—	—
Physical activity	.36	Time only
Blood pressure monitoring	.41	Time only
Body weight monitoring	.32	Time only

Secondary Outcomes

Our study findings showed significant chronological changes in urinary salt concentration between the groups (*P*=.01; Table 4). In the case of food salinity, the salt concentration appeared as thick (0.8%-1.1%) and very thick (1.2%-1.4%), and gradually decreased from the baseline to the study end, with statistically significant changes observed between the 2 groups. Especially, the control group showed an ample drop (*P*<.001), compared to the intervention group (*P*=.001; Figure 3). Conversely, the evaluation of urinary salt excretion from the previous day's salt intake showed inconsistent changes which decreased at the study end. However, mean value of urinary salt excretion did not decrease from 10 g/day at any point.

Systolic BP (SBP) and diastolic BP (DBP) were significantly chronologically decreased in both groups (*P*<.001) over the study period. The mean SBP and DBP dropped between the groups over the study period, more in the intervention group, and the changes were statistically significant (*P*=.04 and *P*=.02, respectively).

In the case of QOL, both groups showed significant improvement at the end of the study (*P*<.001). The intervention group improved more compared to the control group, and a significant difference was observed (*P*<.001) between the groups. However, we did not find any significant interaction between the 2 groups (Table 4).

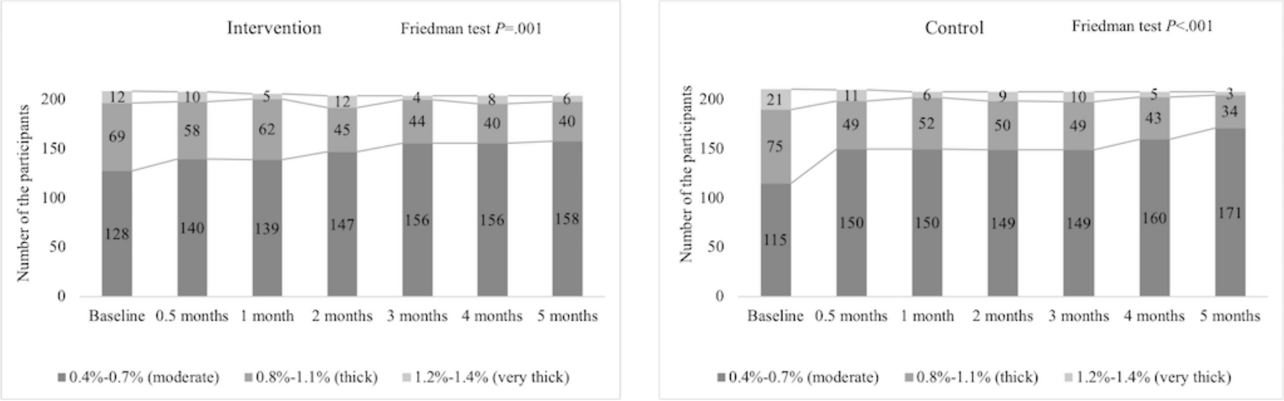
Table 4. Chronological changes in the secondary outcomes between the 2 groups.

Measurement and group	Values							Two-way repeated measures ANOVA					
	Baseline[a]	0.5 months	1 month	2 months	3 months	4 months	5 months	Within groups		Between groups		Interaction	
								F value	P value	F value	P value	F value	P value
Urine salinity (g)[b], mean (SD)													
Intervention group (N=204)	8.04 (2.71)	10.36 (3.11)	10.71 (2.69)	10.85 (2.74)	10.75 (2.80)	10.59 (2.82)	10.18 (2.84)	1.91	.09	2.97	.01	1.62	.15
Control group	8.23 (3.05)	10.93 (2.88)	10.58 (2.94)	10.78 (3.20)	10.75 (3.05)	10.58 (3.20)	10.30 (3.21)						
Systolic blood pressure (mmHg), mean (SD)													
Intervention group (N=204)	136.9 (19.2)	130.0 (16.7)	128.4 (17.1)	127.6 (15.4)	127.9 (17.5)	127.2 (15.7)	125.7 (16.1)	9.95	<.001	4.42	.04	1.7	.12
Control group (N=208)	136.9 (19.3)	132.4 (16.3)	132.3 (16.2)	131.7 (16.1)	132.1 (17.2)	131.6 (16.7)	128.3 (15.4)						
Diastolic blood pressure (mmHg), mean (SD)													
Intervention group (N=204)	89.4 (9.9)	86.3 (9.3)	84.6 (9.0)	84.9 (8.3)	85.3 (9.3)	85.1 (8.4)	84.4 (8.0)	7.71	<.001	5.82	.02	0.38	.89
Control group (N=208)	90.5 (10.4)	87.9 (9.6)	87.0 (9.7)	86.5 (9.5)	87.5 (10.3)	87.4 (10.0)	86.1 (9.3)						
Quality of life (EQ-5D-5L score), mean (SD)													
Intervention group (N=204)	0.70 (0.10)			0.76 (0.11)			0.75 (0.09)	21.58	<.001	17.83	<.001	0.13	.88
Control group (N=208)	0.71 (0.10)			0.76 (0.11)			0.75 (0.09)						

[a] Baseline comparison was tested in each item and no significant differences were observed.

[b] Baseline data of urinary salt excretion was excluded from analysis because it was collected from spot urine, whereas we have collected the overnight urine from the rest of the months.

Figure 3. Food salt concentration changes among the intervention and control groups.



Discussion

Principal Findings

Among the study population, no significant differences were found in baseline demographic characteristics, comorbidities, and medication use as well as behavior as a primary endpoint and BP values, salinity, and QOL as secondary endpoints. Therefore, block randomization was successful. Completion rates of both groups were high and the drop-out rate was only 2%, whereas in previous lifestyle behavior studies the drop-out was 10%-20% [39,40].

In order to make this study more comprehensive, at baseline, we incorporated the RBS and urinary protein investigations as HTN's comorbidities in the analysis. The baseline data estimated the high prevalence of diabetes. Research revealed that more than 80% of comorbidities are associated with HTN where controlling BP is difficult [41]. Therefore, not only behavior change education, but also comprehensive disease management strategies such as dietary change, medication adherence may be required to control HTN. We thus suggest future research to take these issues into consideration.

We expected that SMS text messaging could enhance the effects of behavior changes more than in-person health education alone (10%-12% difference). We thought periodic, weekly SMS text messaging would act as a reminder to change, which can improve self-reported behavior. Nevertheless, the results of this study did not support the expected assumption. The reason for this may perhaps be that relevant, timely, and convenient SMS text messaging can be more effective [42]. As an example, a timely, interactive telephonic behavior change intervention was found to be significantly effective for heart failure patients to prevent re-admission [23].

We supposed that SMS text messaging could be a useful tool for inducing behavior changes in Bangladesh as well as in other low- and middle-income countries due to their fragile health infrastructure [43]. However, we were not able to show the additional effect on in-person health education in the intervention group. In fact, a better improvement was found in the control group compared to the intervention group. The possible reasons may be the following: (1) mHealth technology needs prompt feedback and continued monitoring [44], but in this study we only sent weekly SMS text messages; (2) participants may not have had time to compose the response or may not have had credit on their phones (SMS text messaging needs a small amount of money [credit] to send messages) [33]. Some participants may have reacted as intended but thought that it was not necessary to respond; (3) the conciseness of our SMS text messages may have been difficult to understand (ie, reduce salt intake to 1 teaspoon size); and (4) our SMS text messages were generalized, and more tailored SMS text messages for individuals may have improved the effectiveness of the intervention.

Rather, in-person health education without SMS text messaging was more effective. Our study results also showed that in-person health education had a significant impact on behavior changes among both intervention and control groups. The possible reasons can be, firstly, this study created scopes for Bangladeshi individuals as they have fewer opportunities to receive health education. Health resources are very limited in this country and people do not have a BP machine at home. Besides, they lack opportunity to check BP regularly due to inadequate health check-up systems. As we previously reported [45], people were aware of HTN when they have symptoms. In this study, CHWs have visited home to home and checked BP and food and urine salinity, which was a momentous opportunity for them. Thus, the home health care provided by CHWs is likely a major factor responsible for the improvement of outcomes. Second, the amount of salt intake on the previous day was measured at home using food salinity and urinary salinity devices, which showed the actual status of salt concentration, and participants could connect these data to their BP value.

Our results from the primary outcome showed the significant adherence differences only in salt intake and fruits intake behaviors between the intervention and control groups. Even though the differences were statistically significant, the salt intake change rate was more in the control group from baseline to program endpoint than in the intervention group. Compared to the baseline, almost no participants followed the target goal (<6 g daily salt intake) and this change was noteworthy. Concerning fruits intake behavior, the participants of the intervention group reached the target goal (fruits intake everyday/week) compared with the control group. Moreover, people ate vegetables every day in the rural community, but we did not inquire about the serving amount per day.

Regarding self-reported physical activity behaviors, both groups showed statistically significant differences. Physical activity is beneficial in weight reduction and also improves cardiovascular fitness [46]. Moderate to vigorous physical activity with a goal of 120-150 minutes on at least 5 days in a week should be encouraged [29].

Concerning self-reported BP and weight monitoring behaviors, in the first month (0.5 months and 1 month), participants answered “2 times/month,” because CHWs checked their BP and weight during their follow up. Then they reported “never/week” as none of the participants had facilities to measure these at home. This is possibly due to the cost of BP monitoring and weight-measuring devices, which is too exorbitant for the poor in Bangladesh. Moreover, BP monitoring requires the presence of a trained observer [47,48].

Supporting this self-report, the level of the salt concentration (1.2%-1.4% very thick) in the daily food (liquid) was reduced in both groups, more in the control group. With regard to urine salinity, the urinary salt excretion was reduced at the end of the study; however, it did not reach the recommended goal, and this finding is similar to other studies [49,50]. Combining primary and secondary outcomes, though the primary outcome variables were self-reported, the evidence from secondary outcome variables, which were quantifiable from the physical data, supported the positive changes of primary outcomes. Both SBP and DBP mean values were significantly decreased and SD became narrower in both groups.

As a comprehensive evaluation of this study, QOL of the participants was improved significantly after starting the intervention. This can be attributed to the reception of self-management health education, or the participants might have developed a better understanding of their illness, its effects, greater self-reported adherence, and a better relationship with the CHWs [51]. Thus, the implementation of our intervention created successfully more awareness among the participants.

Strengths and Limitations

In this study, most of the female participants were housewives and thus available at home that allows them to play an influential role than males on behavior changes among the family members [52]. Females also seem to be more engaged, and apparently better-informed decision makers and more familiar in preparing foods, especially in rural areas [53]. This might be another reason for the good results observed on behavior change.

There were some limitations to this study. First, the primary outcome was self-reported by the participants. Second, we did not assess the household income impact as well as other influencing factors that may have had effects on diet among the participants. Third, this study was performed in the context of a small-scale program, and we chose purposive sampling which did not show the generalizability of our results. Fourth, most of the participants were female, and so we cannot generalize our study results to both genders. Fifth, data contamination was highly likely; therefore, the CHWs received verbal consent from the study participants not to disclose any study details with neighbors or other family members.

Conclusions

The results of this study could not show the effectiveness of a combination of SMS text messaging and in-person health education; rather, in-person health education alone had a better impact on behavior changes. Based on the results of study, we suggest that in-person health education requires integration of home health care provision as a major factor for the improvement of outcomes as well as self-management behavior. Besides, it can be recommended that the more relevant and timely interactive SMS text messages can be sent to increase the effectiveness. Finally, the study recommends organizing community awareness meetings to create a “low-salt culture” and educate the family members. Moreover, further research in the diverse population living in different geographical areas is imperative to consolidate or refute our observation.

Acknowledgments

This study was conducted in rural Mirzapur, Bangladesh. We sincerely thank and acknowledge the generous support of the study participants and our CHWs (Ayesha Akter, Sompa Bakali, Tanni Sikder, Shikha Akter, Shanta Islam, Hazera Begum, Afrin Sultana, Hiramon Akter). We are also grateful to Dr Ashir Ahmed, Kyushu University, Japan for providing the Portable Health Clinic device, and Dr Yuko Ito for her continuous support. This study was carried out with funding of the Grants-in-Aid for Scientific Research Program (KAKENHI), Japan (Kiban-B, No. 18H03113).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (V 1.6.1) -Submission/Publication Form.
[PDF File (Adobe PDF File), 1188 KB - jmir_v22i12e19137_app1.pdf]

Multimedia Appendix 2
Awareness development process and content of study outline (referred from reference 27).
[PNG File , 77 KB - jmir_v22i12e19137_app2.png]

Multimedia Appendix 3
Contents of text messages.
[PDF File (Adobe PDF File), 66 KB - jmir_v22i12e19137_app3.pdf]

Multimedia Appendix 4
Behavioral change evaluation questionnaire.
[PDF File (Adobe PDF File), 14 KB - jmir_v22i12e19137_app4.pdf]

Multimedia Appendix 5
Multinomial logit model.
[PDF File (Adobe PDF File), 232 KB - jmir_v22i12e19137_app5.pdf]

Multimedia Appendix 6
Chronological behavior changes.
[PDF File (Adobe PDF File), 99 KB - jmir_v22i12e19137_app6.pdf]

Multimedia Appendix 7
Awareness development process and contents of study outline (referred from reference 27).
[PNG File , 145 KB - jmir_v22i12e19137_app7.png]

References

1. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1659-1724 [FREE Full text] [doi: 10.1016/S0140-6736(16)31679-8] [Medline: 27733284]
2. Kearney PM, Whelton M, Reynolds K, Whelton PK, He J. Worldwide prevalence of hypertension: a systematic review. J Hypertens 2004 Jan;22(1):11-19. [doi: 10.1097/00004872-200401000-00003] [Medline: 15106785]
3. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
4. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet 2005;365(9455):217-223. [doi: 10.1016/S0140-6736(05)17741-1] [Medline: 15652604]
5. O'Brien E. The Lancet Commission on hypertension: Addressing the global burden of raised blood pressure on current and future generations. J Clin Hypertens (Greenwich) 2017 Jun;19(6):564-568 [FREE Full text] [doi: 10.1111/jch.12998] [Medline: 28560771]
6. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. Nat Rev Nephrol 2020 Apr;16(4):223-237. [doi: 10.1038/s41581-019-0244-2] [Medline: 32024986]
7. Alam DS, Chowdhury MAH, Siddiquee AT, Ahmed S, Niessen LW. Awareness and control of hypertension in Bangladesh: follow-up of a hypertensive cohort. BMJ Open 2014 Dec 23;4(12):e004983 [FREE Full text] [doi: 10.1136/bmjopen-2014-004983] [Medline: 25537780]
8. Mendis S. Hypertension: a silent contributor to the global cardiovascular epidemic. Regional Health Forum. 2013. URL: http://origin.searo.who.int/entity/world_health_day/rhf-2013.pdf#page=7 [accessed 2020-11-23]
9. Virdis A, Giannarelli C, Neves MF, Taddei S, Ghiadoni L. Cigarette smoking and hypertension. Curr Pharm Des 2010;16(23):2518-2525. [doi: 10.2174/138161210792062920] [Medline: 20550499]
10. Neupane D, McLachlan CS, Sharma R, Gyawali B, Khanal V, Mishra SR, et al. Prevalence of hypertension in member countries of South Asian Association for Regional Cooperation (SAARC): systematic review and meta-analysis. Medicine (Baltimore) 2014 Sep;93(13):e74 [FREE Full text] [doi: 10.1097/MD.0000000000000074] [Medline: 25233326]
11. Zaman M, Rahman MM, Rahman MR, Bhuiyan M, Karim MN, Chowdhury MA. Prevalence of risk factors for non-communicable diseases in Bangladesh: Results from STEPS survey 2010. Indian J Public Health 2016;60(1):17. [doi: 10.4103/0019-557x.177290]
12. Zaman MM, Choudhury SR, Ahmed J, Khandaker RK, Rouf MA, Malik A. Salt Intake in an Adult Population of Bangladesh. Glob Heart 2017 Sep;12(3):265-266 [FREE Full text] [doi: 10.1016/j.gheart.2016.05.005] [Medline: 27498974]
13. Zaman M, Chowdhury M, Rahman M, Bhuiyan R, Karim M, Islam M, et al. Non-Communicable disease risk factor survey: Bangladesh 2010. 2010. URL: https://apps.who.int/iris/bitstream/handle/10665/279484/9789290223931-eng.pdf [accessed 2020-11-23]
14. Kim Y, Kong KA. Do Hypertensive Individuals Who Are Aware of Their Disease Follow Lifestyle Recommendations Better than Those Who Are Not Aware? PLoS One 2015;10(8):e0136858 [FREE Full text] [doi: 10.1371/journal.pone.0136858] [Medline: 26317336]
15. Anand SS, Yusuf S. Stemming the global tsunami of cardiovascular disease. Lancet 2011 Feb 12;377(9765):529-532. [doi: 10.1016/S0140-6736(10)62346-X] [Medline: 21295845]
16. Tibazarwa K, Ntyintyane L, Sliwa K, Gerntholtz T, Carrington M, Wilkinson D, et al. A time bomb of cardiovascular risk factors in South Africa: results from the Heart of Soweto Study "Heart Awareness Days". Int J Cardiol 2009 Feb 20;132(2):233-239. [doi: 10.1016/j.ijcard.2007.11.067] [Medline: 18237791]
17. Rayner B, Blockman M, Baines D, Trinder Y. A survey of hypertensive practices at two community health centres in Cape Town. S Afr Med J 2007 Apr;97(4):280-284. [Medline: 17446953]
18. Fletcher BR, Hartmann-Boyce J, Hinton L, McManus RJ. The Effect of Self-Monitoring of Blood Pressure on Medication Adherence and Lifestyle Factors: A Systematic Review and Meta-Analysis. Am J Hypertens 2015 Oct;28(10):1209-1221. [doi: 10.1093/ajh/hpv008] [Medline: 25725092]
19. Vargas G, Cajita MI, Whitehouse E, Han H. Use of Short Messaging Service for Hypertension Management: A Systematic Review. J Cardiovasc Nurs 2017;32(3):260-270 [FREE Full text] [doi: 10.1097/JCN.0000000000000336] [Medline: 27111819]
20. Bobrow K, Brennan T, Springer D, Levitt NS, Rayner B, Namane M, et al. Efficacy of a text messaging (SMS) based intervention for adults with hypertension: protocol for the StAR (SMS Text-message Adherence suppoRt trial) randomised controlled trial. BMC Public Health 2014 Jan 11;14:28 [FREE Full text] [doi: 10.1186/1471-2458-14-28] [Medline: 24410738]
21. Schwebel FJ, Larimer ME. Using text message reminders in health care services: A narrative literature review. Internet Interv 2018 Sep;13:82-104 [FREE Full text] [doi: 10.1016/j.invent.2018.06.002] [Medline: 30206523]
22. Márquez Contreras E, de la Figuera von Wichmann M, Gil Guillén V, Ylla-Catalá A, Figueras M, Balaña M, et al. [Effectiveness of an intervention to provide information to patients with hypertension as short text messages and reminders

sent to their mobile phone (HTA-Alert)]. Aten Primaria 2004 Nov 15;34(8):399-405 [FREE Full text] [doi: 10.1016/s0212-6567(04)78922-2] [Medline: 15546536]
23. Mizukawa M, Moriyama M, Yamamoto H, Rahman MM, Naka M, Kitagawa T, et al. Nurse-Led Collaborative Management Using Telemonitoring Improves Quality of Life and Prevention of Rehospitalization in Patients with Heart Failure. Int Heart J 2019 Nov 30;60(6):1293-1302 [FREE Full text] [doi: 10.1536/ihj.19-313] [Medline: 31735786]
24. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
25. Gee ME, Bienek A, Campbell NRC, Bancej CM, Robitaille C, Kaczorowski J, et al. Prevalence of, and barriers to, preventive lifestyle behaviors in hypertension (from a national survey of Canadians with hypertension). Am J Cardiol 2012 Feb 15;109(4):570-575. [doi: 10.1016/j.amjcard.2011.09.051] [Medline: 22154320]
26. Khanam MA, Lindeboom W, Koehlmoos TLP, Alam DS, Niessen L, Milton AH. Hypertension: adherence to treatment in rural Bangladesh--findings from a population-based study. Glob Health Action 2014;7:25028 [FREE Full text] [doi: 10.3402/gha.v7.25028] [Medline: 25361723]
27. Jahan Y, Moriyama M, Rahman MM, Kazawa K, Rahman A, Sayeem Bin Shahid ASM, et al. Increasing Awareness and Use of Mobile Health Technology Among Individuals With Hypertension in a Rural Community of Bangladesh: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2020 Aug 17;9(8):e15523 [FREE Full text] [doi: 10.2196/15523] [Medline: 32804088]
28. Eysenbach G, CONSORT- E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
29. Sacks FM, Svetkey LP, Vollmer WM, Appel LJ, Bray GA, Harsha D, et al. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative Research Group. N Engl J Med 2001 Jan 4;344(1):3-10. [doi: 10.1056/NEJM200101043440101] [Medline: 11136953]
30. Jahan Y, Moriyama M, Rahman MM, Rahman A. Self-monitoring urinary salt excretion device can be used for controlling hypertension for developing countries. Clin Hypertens 2019;25:3 [FREE Full text] [doi: 10.1186/s40885-019-0109-9] [Medline: 30923632]
31. Dirks J, Remuzzi G, Horton S. Diseases of the kidney and the urinary system. In: Jamison DT, Breman JG, Measham AR, editors. Disease Control Priorities in Developing Countries. 2nd edition. Washington DC: The International Bank for Reconstruction and Development/The World Bank; 2006:1-1413.
32. van Hout B, Janssen MF, Feng Y, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15(5):708-715 [FREE Full text] [doi: 10.1016/j.jval.2012.02.008] [Medline: 22867780]
33. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
34. Devlin N, Parkin D. Guidance to users of EQ-5D value sets. In: Szende A, Oppe M, Devlin N, editors. EQ-5D Value Sets. Dordrecht, Switzerland: Springer; 2007:39-52.
35. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. URL: https://www.R-project.org/ [accessed 2020-05-01]
36. Venables W, Ripley B. Modern Applied Statistics with S. 4th edition. New York: Springer; 2002:1-497.
37. Wickham H, Francois R, Henry LK. dplyr: a grammar of data manipulation. R package version 0. 2020. URL: https://CRAN.R-project.org/package=dplyr [accessed 2020-05-02]
38. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer; 2016.
39. National Research Council (US) Panel on Race, Ethnicity, and Health in Later Life. In: Anderson NB, Bulatao RA, Cohen B, editors. Critical Perspectives on Racial and Ethnic Differences in Health in Late Life. Washington, DC: National Academies Press; 2004.
40. Cutler D. Behavioral Health Interventions: What Works and Why?. Washington, DC: National Academies Press; 2004:1-24.
41. Steigerwalt S. Management of Hypertension in Diabetic Patients With Chronic Kidney Disease. Diabetes Spectrum 2008 Jan 01;21(1):30-36. [doi: 10.2337/diaspect.21.1.30]
42. Hacking D, Haricharan HJ, Brittain K, Lau YK, Cassidy T, Heap M. Hypertension Health Promotion via Text Messaging at a Community Health Center in South Africa: A Mixed Methods Study. JMIR Mhealth Uhealth 2016 Mar 10;4(1):e22 [FREE Full text] [doi: 10.2196/mhealth.4569] [Medline: 26964505]
43. Islam A. Health System in Bangladesh: Challenges and Opportunities. AJHR 2014;2(6):366-374. [doi: 10.11648/j.ajhr.20140206.18]
44. Samples C, Ni Z, Shaw RJ. Nursing and mHealth. International Journal of Nursing Sciences 2014 Dec;1(4):330-333. [doi: 10.1016/j.ijnss.2014.08.002]
45. Jahan Y, Moriyama M, Rahman MM, Kazawa K, Mizukawa M, Rahman A, et al. Disease perception and experiences among rural Bangladeshi hypertensive women: A qualitative approach. Health Promot Perspect 2020;10(1):66-73 [FREE Full text] [doi: 10.15171/hpp.2020.11] [Medline: 32104659]

46. Farpour-Lambert NJ, Aggoun Y, Marchand LM, Martin XE, Herrmann FR, Beghetti M. Physical activity reduces systemic blood pressure and improves early markers of atherosclerosis in pre-pubertal obese children. J Am Coll Cardiol 2009 Dec 15;54(25):2396-2406 [FREE Full text] [doi: 10.1016/j.jacc.2009.08.030] [Medline: 20082930]
47. Begum HA, Perveen R, Chakma E, Dewan L, Afroze RS, Tangen D. The challenges of geographical inclusive education in rural Bangladesh. International Journal of Inclusive Education 2018 Sep 02;23(1):7-22. [doi: 10.1080/13603116.2018.1514729]
48. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Joint National Committee on Prevention, Detection, Evaluation,Treatment of High Blood Pressure. National Heart, Lung,Blood Institute, National High Blood Pressure Education Program Coordinating Committee. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension 2003 Dec;42(6):1206-1252. [doi: 10.1161/01.HYP.0000107251.49515.c2] [Medline: 14656957]
49. Ha SK. Dietary salt intake and hypertension. Electrolyte Blood Press 2014 Jun;12(1):7-18 [FREE Full text] [doi: 10.5049/EBP.2014.12.1.7] [Medline: 25061468]
50. Ndanuko R, Tapsell L, Charlton K, Neale E, Batterham M. Dietary Patterns and Blood Pressure in Adults: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Adv Nutr 2016;7(1):89. [doi: 10.3945/an.115.009753] [Medline: 26773016]
51. Aghajani M, Mirbagher Ajorpaz N, Kafaei Atrian M, Raofi Z, Abedi F, Naeimi Vartoni S, et al. Effect of self - care education on quality of life in patients with primary hypertension: comparing lecture and educational package. Nurs Midwifery Stud 2013 Dec;2(4):71-76 [FREE Full text] [doi: 10.5812/nms.11655] [Medline: 25414882]
52. Hernandez EM, Margolis R, Hummer RA. Educational and Gender Differences in Health Behavior Changes After a Gateway Diagnosis. J Aging Health 2018 Mar;30(3):342-364 [FREE Full text] [doi: 10.1177/0898264316678756] [Medline: 27940641]
53. Ek S. Gender differences in health information behaviour: a Finnish population-based survey. Health Promot Int 2015 Sep;30(3):736-745 [FREE Full text] [doi: 10.1093/heapro/dat063] [Medline: 23985248]

Abbreviations

BP: blood pressure
CHWs: community health workers
DASH: dietary approaches to stop hypertension
HTN: hypertension
NCDs: noncommunicable diseases
QOL: quality of life
RBS: random blood sugar

Edited by G Eysenbach; submitted 05.04.20; peer-reviewed by F Schwebel; comments to author 29.04.20; revised version received 23.05.20; accepted 26.10.20; published 07.12.20.

Please cite as:
Jahan Y, Rahman MM, Faruque ASG, Chisti MJ, Kazawa K, Matsuyama R, Moriyama M
Awareness Development and Usage of Mobile Health Technology Among Individuals With Hypertension in a Rural Community of Bangladesh: Randomized Controlled Trial
J Med Internet Res 2020;22(12):e19137
URL: https://www.jmir.org/2020/12/e19137
doi:10.2196/19137
PMID:33284129

Original Paper

A Mobile Social Networking App for Weight Management and Physical Activity Promotion: Results From an Experimental Mixed Methods Study

Liliana Laranjo[1,2], MPH, MD, PhD; Juan C Quiroz[1,3*], PhD; Huong Ly Tong[1*], MRes; Maria Arevalo Bazalar[1], MPH; Enrico Coiera[1], MBBA, PhD

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

[2]Westmead Applied Research Centre, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[3]Centre for Big Data Research in Health, University of New South Wales, Sydney, Australia

*these authors contributed equally

Corresponding Author:
Liliana Laranjo, MPH, MD, PhD
Australian Institute of Health Innovation
Macquarie University
Level 6, 75 Talavera road
Sydney, 2113
Australia
Phone: 61 413461852
Email: liliana.laranjo@mq.edu.au

Abstract

Background: Smartphone apps, fitness trackers, and online social networks have shown promise in weight management and physical activity interventions. However, there are knowledge gaps in identifying the most effective and engaging interventions and intervention features preferred by their users.

Objective: This 6-month pilot study on a social networking mobile app connected to wireless weight and activity tracking devices has 2 main aims: to evaluate changes in BMI, weight, and physical activity levels in users from different BMI categories and to assess user perspectives on the intervention, particularly on social comparison and automated self-monitoring and feedback features.

Methods: This was a mixed methods study involving a one-arm, pre-post quasi-experimental pilot with postintervention interviews and focus groups. Healthy young adults used a social networking mobile app intervention integrated with wireless tracking devices (a weight scale and a physical activity tracker) for 6 months. Quantitative results were analyzed separately for 2 groups—underweight-normal and overweight-obese BMI—using *t* tests and Wilcoxon sum rank, Wilcoxon signed rank, and chi-square tests. Weekly BMI change in participants was explored using linear mixed effects analysis. Interviews and focus groups were analyzed inductively using thematic analysis.

Results: In total, 55 participants were recruited (mean age of 23.6, SD 4.6 years; 28 women) and 45 returned for the final session (n=45, 82% retention rate). There were no differences in BMI from baseline to postintervention (6 months) and between the 2 BMI groups. However, at 4 weeks, participants' BMI decreased by 0.34 kg/m^2 ($P<.001$), with a loss of 0.86 kg/m^2 in the overweight-obese group ($P=.01$). Participants in the overweight-obese group used the app significantly less compared with individuals in the underweight-normal BMI group, as they mentioned negative feelings and demotivation from social comparison, particularly from upward comparison with fitter people. Participants in the underweight-normal BMI group were avid users of the app's self-monitoring and feedback ($P=.02$) and social ($P=.04$) features compared with those in the overweight-obese group, and they significantly increased their daily step count over the 6-month study duration by an average of 2292 steps (95% CI 898-3370; $P<.001$). Most participants mentioned a desire for a more personalized intervention.

Conclusions: This study shows the effects of different interventions on participants from higher and lower BMI groups and different perspectives regarding the intervention, particularly with respect to its social features. Participants in the overweight-obese group did not sustain a short-term decrease in their BMI and mentioned negative emotions from app use, while participants in the underweight-normal BMI group used the app more frequently and significantly increased their daily step count. These differences highlight the importance of intervention personalization. Future research should explore the role of personalized features to help overcome personal barriers and better match individual preferences and needs.

(J Med Internet Res 2020;22(12):e19991) doi:10.2196/19991

KEYWORDS

mobile apps; fitness trackers; exercise; social networking; body weight maintenance; mobile phone

Introduction

Background

Obesity and physical inactivity are major societal challenges and significant contributors to the global burden of disease and health care costs [1-3]. Globally, excess body weight and obesity account for approximately 4 million deaths in a year and 120 million disability-adjusted life years [1]. Effective weight management interventions usually involve a combination of behavioral strategies focusing on dietary changes and physical activity [4]. Notably, physical activity is also independently associated with better health outcomes in a dose-response manner [5], however, more than one-fourth of adults globally are insufficiently active [6]. Promoting physical activity and weight management (ie, preventing weight gain to unhealthy levels and promoting weight loss in individuals with excess weight) are important behavioral strategies for better health outcomes.

Addressing obesity and physical inactivity requires a combination of environmental approaches and effective behavior change interventions that can be easily disseminated. Apps and fitness trackers are becoming pervasive in the daily lives of people, with smartphone ownership surpassing three-fourth of the population and activity trackers being used by one-third of adults in the United States and United Kingdom [7,8]. These mobile technologies enable the automation of effective behavior change techniques—weight and physical activity self-monitoring and feedback [9-12]—showing potential in weight management and physical activity interventions [13,14].

Although mobile technologies can facilitate self-monitoring and feedback, behavior change is also heavily regulated by social processes. Online social networks—platforms that allow individuals to create their own personal profile and build a network of connections with other users—can facilitate behavior change [15-18] and have shown potential in weight management and physical activity interventions [19-22]. Online social networks enable the delivery of several social functions, including support and social comparison, which have been associated with increased physical activity [23-25] and greater weight loss in weight management interventions [26,27]. However, user preferences regarding social features seem to be mixed: for some users, these features promote engagement with the intervention, whereas for other users, they are less enjoyable or even disliked [20,28]. Key questions remain as to which intervention features and social network characteristics are more effective and engaging for users in weight management and physical activity promotion [15,18].

Aims

This 6-month pilot study of a social networking mobile app connected to wireless weight and activity tracking devices has 2 main aims: (1) to evaluate changes in BMI, weight, and physical activity levels in users from different BMI categories and (2) to assess user perspectives on the intervention, particularly its online social networking component and automated self-monitoring and feedback features.

Methods

Study Design

This was a mixed methods study involving a pre-post quasi-experimental pilot with one arm, where participants used a social networking mobile app intervention integrated with wireless tracking devices (a weight scale and a physical activity tracker) for 6 months [29]. Ethics approval was granted by the Human Research Ethics Committee for Medical Sciences of Macquarie University. Reporting follows the mobile health (mHealth) evidence reporting and assessment (mERA) checklist [30], the transparent reporting of evaluations with nonrandomized designs (TREND) statement for reporting of nonrandomized evaluations of behavioral and public health interventions [31], the COnsolidated Criteria for Reporting Qualitative Research (COREQ) checklist [32], and guidelines for good reporting of mixed methods studies [33] (Multimedia Appendices 1-3).

Study Sample and Recruitment

Eligible study participants were consenting healthy adults who were able to speak, write, and read English, were between 19 and 35 years of age, owned a mobile phone (iOS or Android) with internet access, and had Wi-Fi connection at home. Exclusion criteria were pregnancy or breastfeeding, BMI below 17 kg/m^2, prior history of eating disorders, and having diabetes or other comorbid conditions that could have impacted study participation.

Students and staff from Macquarie University (Sydney, Australia) were recruited via posters and flyers distributed around campus as well as via a Facebook post on the university's page (Multimedia Appendix 4). Recruitment was completed in April 2017 and followed a purposive convenience sampling technique to test the intervention in a diverse BMI sample. To enable a comparable number of individuals in lower and higher BMI groups (at least 20 individuals in each of those 2 groups), recruitment of individuals in the normal BMI range was stopped earlier, whereas recruitment for the higher BMI groups continued. A sample size of at least 40 individuals was pragmatically chosen to enable pilot testing of the intervention based on available funding. At baseline, participants were invited to attend the initial study session at a research center on campus, where they signed the consent form and filled in a questionnaire about their demographic characteristics. A complete description of study procedures and interventions can be found in published papers [29,34,35].

Intervention

The intervention consisted of a mobile app (fit.healthy.me) [29,34,35] designed by the research team, which was integrated with 2 wireless devices (Fitbit Aria weight scale and Fitbit Flex 2 physical activity tracker, connected via the Fitbit app). The goal of the intervention was to promote physical activity and support weight management in users of any physical activity level or BMI group. In fit.healthy.me, the participants could compare their step count and weight with other users, in table and graphical formats (Multimedia Appendix 5). Furthermore, users were able to interact and provide social support to each other through the use of messaging and a social forum, as well as *follow* particular *buddies* with whom they identified more closely.

The intervention allowed for the delivery of several behavior change techniques [11]: self-monitoring and feedback on behavior (daily number of steps), self-monitoring and feedback on weight and BMI, instructions on how to perform the behavior, social support, and social comparison. Prompts and cues (text messages and emails every 2 weeks) were used to promote engagement with the intervention. Goal setting was not incorporated in fit.healthy.me but it was a core component of the Fitbit devices (eg, the activity tracker had a predefined daily step goal of 10,000, which could be modified by users); goal setting was neither promoted nor discouraged by researchers.

Quantitative Data Collection and Analysis

Weight and BMI

The primary outcome was the difference in the average BMI between 6 months and baseline. Body weight was measured with a Fitbit Aria scale to the nearest 0.1 kg in light clothing without shoes, before and after the intervention period. Height was measured using a wall-mounted stadiometer to the nearest 0.1 cm. BMI was calculated as weight (kg)/height^2 (m^2). Weight and BMI were also measured by participants at several time points throughout the 6 months (participants were asked to weigh themselves daily using the Fitbit Aria scale provided for research purposes, to enable the comprehensive testing of the integration between the fit.healthy.me app and the wireless scale, and early detection of potential problems).

Daily Step Count

The daily step count was measured using the Fitbit Flex 2. To establish a baseline average daily step count, participants did not have access to the intervention app for the first week after enrollment. Participants were considered to have a valid step count if they wore the Fitbit for at least 10 hours on any given day. Wear time was calculated by subtracting nonwear time from 24 hours; nonwear time was defined when step counts over a period of at least 60 continuous minutes were zero (allowing for counts of less than 100 for a maximum of 2 min within that period) [35].

Engagement Measures

Retention was defined as attendance at the 6-month final session; participants who attended the final sessions were considered *completers*. For the fit.healthy.me app, engagement was measured by number of app sessions and frequency of usage of app features (ie, the number of times participants used each feature). One app session was defined as any activity occurring in the app until the user logged off or when 10 consecutive minutes of inactivity were reached. A participant was considered to have used a social feature if they clicked on any of *My team*, *Social forum*, and *Private messages* features.

Data Analysis

Quantitative results were analyzed separately for 2 groups: underweight-normal BMI (18-24.99 kg/m^2) and overweight-obese (≥25 kg/m^2). Missing weight and step count data were imputed using the last observation carried forward (last measurement obtained from the Fitbit devices). Independent two-sample *t* tests were used for normally distributed continuous variables. For nonnormal data, the Wilcoxon sum rank test was used (Wilcoxon signed rank test for paired within-group comparisons). Chi-square tests were used for categorical data.

We performed an exploratory linear mixed effects analysis of the weekly BMI change of each participant. The dependent variable was BMI change. We used intercepts, sex, age, and weight baseline as fixed effects. As random effects, we had intercepts for subjects as well as by-subject random slopes for the effect of time. We used 6 months of data, with the origin of the time variable (week) set at week 4 (posthoc decision), for 2 main reasons: the likelihood of observing weight changes due to the intervention would be higher at 4 weeks rather than before (given that weight loss increases with intervention duration [36]) and the amount of missing data was minimal during the first month of the intervention, allowing for a more robust model [37]. *P* values were obtained by using likelihood ratio tests of the full model with the effect in question against the model without it. Data were analyzed using R version 3.5.0 and Ime4 [38]. The significance level for all statistical tests was set at *P*<.05, two tailed, and 95% CIs were calculated where applicable.

Qualitative Data Collection and Analysis

An interview guide [34] was pilot tested before study commencement. In the final 6-month session, 2 researchers with expertise in qualitative methods conducted individual interviews and focus groups with participants to understand their perspectives on the advantages and disadvantages of the intervention, until data saturation was reached. The interviews allowed us to understand individual perspectives, and the focus groups aimed to explore group differences and similarities. Field notes were taken throughout the interviews and focus groups. Interviews and focus groups were audio-recorded and transcribed verbatim.

All data were imported and managed in Nvivo 11 (QSR International). Data were analyzed using thematic analysis, where transcripts and field notes were read to identify and code common ideas and patterns emerging from the data [39]. Through constant comparison, codes and concepts were clustered together to form subthemes, and further abstracted to originate themes, which were then reviewed and refined [40]. The integration of results was done after quantitative and qualitative analyses were conducted through embedding of the data. Integration is presented throughout the *Discussion* section.

Results

Sample Characteristics

A total of 55 participants were recruited, with a mean age of 23.6 (SD 4.6) years; 51% (n=28) were female (Table 1). Most (n=24, 44%) participants had a normal BMI, 27% (n=15) were overweight, 24% (n=13) were obese, and 5% (n=3) were underweight. The mean step count per day at baseline was 9937 (SD 3527). Of the 55 recruited participants, 45 returned for the final session (study completers)—the retention rate was 82%. There were no statistically significant differences in baseline characteristics between enrolled participants and study completers.

Table 1. Baseline characteristics of the study sample, according to BMI categories.

Variable	BMI categories (kg/m^2)[a]				Enrolled participants (N=55)	Study completers (n=45)
	18-18.49 (n=3)	18.5-24.99 (n=24)	25-29.99 (n=15)	≥30 (n=13)		
Age (years), mean (SD)	22.2 (3.3)	22.2 (3.6)	25.6 (5.6)	24.1 (4.8)	23.58 (4.6)	24.2 (4.7)
Female, n (%)	2 (67)	15 (63)	4 (25)	7(25)	28 (51)	22 (49)
Weight (kg), mean (SD)	54.3 (5.0)	65.6 (7.9)	84.4 (8.0)	107.1 (22.7)	78.1 (22.3)	77.8 (21.2)
BMI (kg/m^2), mean (SD)	19.4 (1.3)	22.6 (2.3)	27.5 (1.6)	36.9 (5.5)	26.5 (6.8)	26.7 (6.5)
Steps per day[b], mean (SD)	8203 (2824)	9619 (1720)	12,128 (3820)	8912 (3345)	9937 (3527)	9946 (3656)

[a] According to the World Health Organization, a BMI of <18.5 is classified as underweight, 18.5-24.9 is normal, 25-29.9 is preobese, and ≥30 is obese.

[b] Analysis of variance: *P*=.03.

Quantitative Findings

Weight and BMI

There were no statistically significant differences in BMI from baseline to postintervention (6 months) and between the underweight-normal and overweight-obese groups (Table 2; Figure 1). Linear mixed effects analyses of the weekly BMI change of each participant are shown in Table 3. Age, pre-post step difference, and scale usage did not have an effect on BMI difference in any of the 3 groups (all participants; underweight-normal BMI; and overweight-obese; Multimedia Appendix 6). A model with sex and baseline weight as fixed effects was not statistically different from the model presented in Table 3, with just sex as a fixed effect (parameters for the model with sex and baseline weight as fixed effects are included in Multimedia Appendix 7).

At 4 weeks of the intervention, participants' BMI decreased by 0.34 kg/m^2 (0.86 kg/m^2 in the overweight-obese group). Over the 6-month study, men showed a 0.32 increase in BMI relative to women during the intervention (Table 3). The variability in the rate of BMI change across all participants over the duration of the intervention was low (≤0.004), but it was 4 times higher in the overweight-obese group (BMI change variance of 0.004) compared with the underweight-normal group (BMI change variance of 0.001).

Table 2. Differences in characteristics of underweight-normal BMI and overweight-obese participants.

Variable	Underweight-normal BMI[a] (n=27), mean (SD)	Overweight-obese (n=28), mean (SD)	*P* value (95% CI)[b]
Baseline steps per day	9094 (2916.2)	10,749 (3910.8)	.10 (–3299 to 255)
Pre-post step difference	2292 (3520.4)[c]	–213.7 (4023.4)[d]	*.049*[e] (4.6 to 3891)
Pre-post BMI difference	0.13 (0.68)	–0.23 (1.58)	.30 (–0.14 to 0.82)
Scale use	111.9 (65.4)	119.4 (122)	.60 (–36 to 60)
App sessions[f]	23.5 (23.8)	15.8 (20)	.08 (–1 to 13)
App use frequency	1894 (1653.5)	1384.3 (1192.2)	.30 (–313 to 1100)
My measures use	59.2 (54.8)	29.7 (35)	*.02* (3 to 45)
Social features use[g]	213.3 (208.5)	110.1 (139)	*.04* (0 to 167)
System usability scale score[h]	60.8 (20.7)	59.4 (18.1)	.34 (–5 to 15)

[a] Using baseline measurements; normal BMI: BMI 17-24.9 kg/m^2, overweight-obese: BMI of ≥25.

[b] Wilcoxon sum rank test.

[c] Within-group (Wilcoxon sign rank test): *P*<.001; 95% CI 898 to 3370.5.

[d] Within-group: *P*=.73; 95% CI 1365.7 to 1486.4.

[e] Italics denotes statistical significance.

[f] One app session was defined as any activity occurring in the app until the user logs off or when 10 min of inactivity are reached.

[g] Social features include *My team*, *Social forum*, *Private messages*.

[h] Only study completers (ie, participants who returned to the final session) completed the system usability scale (n=45).

Figure 1. Box plots of pre-post BMI difference, pre-post step count difference, My measures use frequency, and social features use frequency.

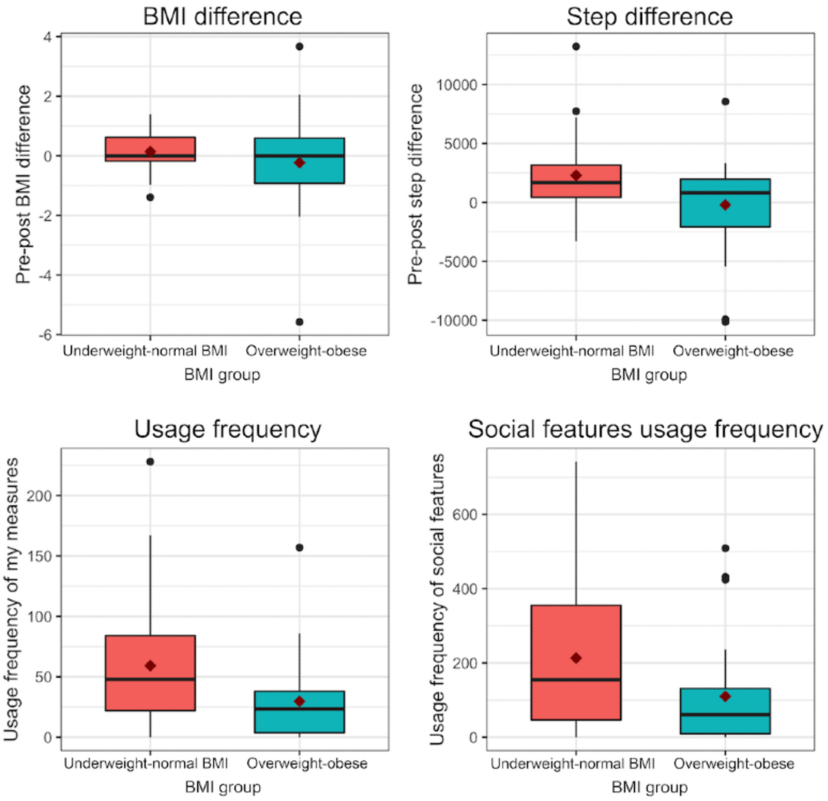


Table 3. Linear mixed effects analysis of the weekly BMI change of each participant.

Predictors[a]	BMI											
	All participants				Underweight-normal				Overweight-obese			
	Estimate	SE	95% CI	*P* value	Estimate	SE	CI	*P* value	Estimate	SE	95% CI	*P* value
Intercept	–0.34	0.10	(–0.5 to –0.1)	*<.001* [b]	–0.23	0.28	(–0.8 to 0.3)	.41	–0.86	0.33	(–1.5 to –0.2)	*.01*
Sex	0.32	0.14	(0.05 to 0.6)	*.02*	0.17	0.16	(–0.2 to 0.5)	.31	0.34	0.22	(–0.1 to 0.8)	.12
Weekly BMI change	–0.01	0.01	(–0.02 to 0.01)	.28	0.01	0.01	(–0.01 to 0.02)	.22	–0.02	0.01	–0.05 to 0.0)	.06

[a] Random effects: intercept variance 0.27 (all participants), 0.15 (underweight-normal), 0.34 (overweight-obese); weekly BMI change variance 0.002 (all participants), 0.001 (underweight-normal), 0.004 (overweight-obese); covariance intercept–weekly BMI change 0.37 (all participants), 0.06 (underweight-normal), 0.37 (overweight-obese).

[b] Italics denotes statistical significance.

Daily Step Count

The underweight-normal group increased their daily step count over the 6-month study duration by an average of 2292 steps (SD 3520.4; 95% CI 898-3370; *P*<.001), whereas the overweight-obese group did not show statistically significant changes (Table 2; Figure 1).

Engagement Measures

App usage frequency was significantly higher in the underweight-normal BMI group: over half of the participants in this BMI group used the *My measures* feature at least 50 times and used social features 150 times or more over the 6-month study duration (Table 2; Figure 1).

Qualitative Findings

We conducted 32 individual interviews and 5 focus groups with 13 participants (20-45 min for each interview or focus group). Themes and subthemes did not differ between interviews and focus groups and consisted of (1) social comparison and networking (subthemes: social comparison, digital people watching and data sharing, social interactions, and negative aspects of social interactions), (2) self-monitoring (subthemes: self-monitoring of weight and BMI, self-monitoring of steps, and other fitness-related measures), and (3) Personalization and gamification. Social comparison and self-monitoring were seen as very distinct features by users, each with its positive and negative aspects, which in turn influenced motivation to use the app and engagement with the intervention. Personalization and gamification were commonly mentioned as desirable features to promote long-term engagement.

Social Comparison and Networking

Social Comparison

Most participants mentioned that social comparison of weight, BMI, or step count can be pointless if participants' characteristics, goals, and lifestyle are not known (Textbox 1, quotes 1-3). However, most participants mentioned being motivated by the competition aspects enabled by social comparison, especially with regard to physical activity (Textbox 1, quote 4). The preferred type of comparison varied between individuals. Several participants in the underweight-normal category mentioned a preference for upward comparison in terms of fitness level, where they enjoyed comparing themselves with more active people in the lower BMI ranges (Textbox 1, quotes 5 and 6), whereas some participants in the overweight and obese categories indicated an inclination for comparison with individuals in similar or higher BMI categories to avoid feeling demotivated (Textbox 1, quotes 7-9).

Textbox 1. Illustrative quotes for social aspects related to weight management and physical activity.

Social comparison:

- Common opinions
  - 1. Comparison never helped me at all. It never motivated me to add them to my friend list or anything, or compare my data with them regularly. Because I don't know them first of all, I don't know what their goal is, I don't know what their existing lifestyle is like or anything of that sort. I didn't know how to compare myself with them. (Male, 27, normal BMI)
  - 2. Feeling (...) disconnected from the numbers. (...) The numbers meant nothing really. [About the comparison of step counts and BMI with other participants] (Male, 20, high BMI)
  - 3. If everyone has the same goal (...) then it's beneficial, then the competitive aspect comes into play. (Male, 20, normal BMI)
  - 4. There were a lot of challenges that my friends also kept adding, like in [Fitbit] (...) with their friends and with my friends combined. So, it became like (...) 25, 30 people in one challenge. (...) Yeah, it was pretty good actually because it kept me motivated (...) the competition thing. (Male, 27, normal BMI)
- Upward comparison
  - 5. I prefer to just (...) compare myself with the more athletic people because I see myself as more of an athlete than not. (Male, 20, normal BMI)
  - 6. Somebody who you want to look up to and how they're using the device to get their goals—that will be a good value-add in my life. (Male, 27, normal BMI)
- Downward comparison
  - 7. I was looking at people who are similar and started off in a similar position, and then I just tried to keep track of who did - like how much walking they did and how their weight was going and so on. (Male, 22, high BMI)
  - 8. I reckon if you had groups, if you had all the fitness people together and then all of the average people together and then all the overweight people, it would feel more - you wouldn't feel as bad because you're finding out there are other people like you. Even if all of your group BMI is in the high range, you would - you'd feel like that there are other people like you who are trying. (Female, 20, high BMI)
  - 9. I went through [all the participants] and got all the ones with similar BMIs and that way at least my buddies were what I would consider similar, like (...) female and same BMI. (Female, 24, high BMI)

Digital *people watching* and data sharing

- 10. I like how we get to compare, and I like how other people don't get to see that I'm comparing against them. (...) So I don't look like a stalker. (Female, 20, normal BMI)
- 11. I guess there's not enough information in the app itself to kind of identify that person. I'm more than happy for other people to find out (...) how many steps I take and how much I weigh and how tall I am. Mainly because I don’t (...) have [a] close connection with them. If I did, I'd probably (...) be reserved in exposing some of that information. (Male, 30, high BMI)

Social interactions

- 12. For me it's the face-to-face and seeing someone and having to be accountable like that. (Female, 25, high BMI)
- 13. If you see that someone may be exercising around the same time as you, it might be you could turn it into a social thing where maybe you go to the gym or the park together. (Male, 22, high BMI)

Negative aspects of social interventions

- 14. When I realised I wasn't doing well against other people (...) I [wouldn’t] check it as much. (Female, 24, high BMI)
- 15. it can be a little bit demotivating when other people are not (...) taking this seriously (Female, 24, normal BMI)

Digital People Watching and Data Sharing

Several people mentioned that they enjoyed looking at other peoples’ measures without them knowing, a digital form of *people watching* enabled by the social network component of the app (Textbox 1, quote 10). Interestingly, sharing their own data, particularly weight and BMI, was disliked by some participants in the higher BMI categories, who mentioned a preference to remain anonymous in the network (Textbox 1, quote 11).

Social Interactions

Most participants highlighted the need to have real-world social connections for social support and accountability during the intervention. The desire for face-to-face support was particularly mentioned by people in the higher BMI categories, who indicated a need to have people to exercise with or a personal trainer to hold them accountable (Textbox 1, quotes 12 and 13). Participants in the normal BMI category frequently mentioned that they would have liked to have been able to invite their friends and family to the intervention, with some of them having even bought Fitbit trackers for family members to encourage their physical activity efforts (Textbox 1, quote 4).

Negative Aspects of Social Interventions

Two main negative aspects were mentioned with regard to social interventions. First, social comparison was seen by many participants as potentially leading to demotivation and negative emotions such as frustration, especially in the case of upward comparison in individuals with higher BMI, that is, comparison with a higher standard (lower BMI or more active individuals; Textbox 1, quote 14) [41]. Second, even highly motivated individuals could be negatively impacted by the lack of motivation or engagement from other individuals, such as participants not showing steps or weight data for periods of time (Textbox 1, quote 15). Both of these social causes for demotivation and negative emotions (upward comparison and lack of participation from other individuals) were mentioned by participants as leading to their lower engagement with the intervention.

Self-Monitoring

Self-Monitoring of Weight and BMI

Self-monitoring of weight and BMI was seen as *pointless* by several people in the normal weight range, although a few individuals mentioned liking the increased awareness of variations in their weight, especially when trying to gain muscle mass (Textbox 2, quotes 1 and 2). Some individuals in the higher BMI group saw the importance of weight monitoring and the benefits of using a wireless scale. However, individuals with both normal and higher BMI indicated a potential for negative emotions associated with self-monitoring, such as during periods of weight gain (Textbox 2, quotes 3 and 4).

Textbox 2. Illustrative quotes for self-monitoring of weight, BMI, and physical activity.

Self-monitoring of weight and BMI:

- 1. the weight (...) doesn't really matter to me that much. (Female, 24, normal BMI)
- 2. I think when I started, I was 58.8 and now I'm 59.8. (...) I didn't want to lose weight (...). I wanted to gain muscle, so hopefully that one kilo is muscle. (Female, 20, normal BMI)
- 3. I didn't like the weight [monitoring]. (...) I didn't like looking at it there every single day, but I would check it. (Female, 31, high BMI)
- 4. When I was weighing myself (...) [and I felt] like I was gaining weight (...) I felt a bit stressed. (...) I was aware that I was gaining weight and then at times I was trying to eat less but then I ended up eating more chocolate, as in binge eating. (Female, 20, normal BMI)

Self-monitoring of steps and other fitness-related measures:

- 5. I would just open the app just to check (...) how many steps I [had] done. Because my step goal was about 10,000 steps and I would do that almost every day. (Male, 27, normal BMI)
- 6. At first, I was motivated and I was going for exercises because of the Fitbit just to achieve the goals. Then afterwards I lost interest. (Female, 20, high BMI)
- 7. Definitely at the beginning I was doing more, but then - I don't know - halfway through I just [got] bored of it. I needed something else. (...) I mean, life's busy, so you just forget about it. (Female, 22, high BMI)
- 8. In the beginning, yeah, I found it really cool. I could track my steps. I tried to complete my goal every day. (...) I was really, really motivated, I was tracking everything. But then I got bored of it really quickly and that's how I started to lose interest. (Female, 20, normal BMI)
- 9. I didn't like to be reminded if I didn't achieve a goal. I feel like there's so much in our lives that we [have to] do that (...) getting a reminder that you haven't achieved whatever your goal [was] I found that I didn't like that at all, to be honest. I found that to be demotivating. I just didn't want to know about it. (Female, 31, high BMI)
- 10. When I got the highest [number of] steps, I was pretty impressed. I took a screenshot. I was bragging about it. (...) Because that wasn't to do with weight or BMI or whatever, it was more of an equal playing field, I feel. (...) (Female, 20, high BMI)
- 11. I play basketball, so I [want to] compare myself to my rivals. Someone who plays in a similar position and skillset as me. But if it's just number of steps, how much walking I did on the day, I don't really want to compare. (...) Step count is just walking. I don't find walking a competitive thing. (Male, 31, high BMI)
- 12. [About the importance of measuring heart rate] Because with all the exercise, it changes up a lot - I wanted to see what my resting heart rate was.

Self-Monitoring of Steps and Other Fitness-Related Measures

Self-monitoring of steps was particularly useful for individuals who had a specific daily step count goal, independent of BMI group (Textbox 2, quote 5). Many participants mentioned that having a goal was a necessary but not sufficient condition for engagement. Specifically, while it was motivating in the beginning to try to achieve new goals, the novelty effect often wore off, with participants going back to previous patterns of physical activity (Textbox 2, quotes 6-8). During periods of decreased physical activity and lower step counts, several individuals mentioned staying away from checking the mobile app to avoid negative emotions and feelings of guilt (Textbox 2, quote 9).

Some participants in the higher BMI groups seemed to prefer self-monitoring of steps to self-monitoring of weight or BMI, owing to an increased sense of control over changes and a higher ability to progress and achieve goals (Textbox 2, quote 10).

Self-monitoring of steps was seen as meaningless by other participants involved in types of physical activity where step counts were either irrelevant or not a good reflection of the effort or performance in the activity (eg, swimming, basketball; Textbox 2, quote 11). In addition, for a few individuals, wearing the Fitbit tracker was not possible while performing their sport (eg, rugby), decreasing their interest in the self-monitoring of steps. Users interested in increasing their fitness levels or muscle mass indicated a preference for other measurements in terms of self-monitoring (eg, heart rate, body fat percentage), instead of weight, BMI, or step counts (Textbox 2, quote 12).

Personalization and Gamification

Most participants mentioned similar preferences and desires for features, namely personalization and gamification. In particular, most people indicated the desire to have a higher degree of personalization in the app, in terms of the features, content, and feedback provided (Textbox 3, quotes 1-4). In addition, several participants highlighted the powerful role of gamification in increasing and sustaining motivation and engagement with the intervention (Textbox 3, quotes 5 and 6). Gamified features were considered important in both individual aspects related to weight management and physical activity (eg, badges for goal achievements) and in social aspects such as social comparison or competition (eg, leaderboards).

Textbox 3. Illustrative quotes for preferred and desired features—personalization and gamification.

Preferred and desired features:

- Personalization
    - 1. You think about gym training sessions, you can have private sessions, you can have small group sessions or you can have a class session and you choose which one is best for you. The same [should happen] with the app. (Female, 20, normal BMI)
    - 2. I personally thought [the app] would give me recommendations on easier [exercises]. (...) Tailor it to me [according to the reaction to previous exercises] (Female, 20, high BMI)
    - 3. [Having health information] would be good, but it has to be personalized or customized to me, (...) my body type, (...) not like a general advice like [what is] BMI etc. (...) A lot of people can read about general information; but if it's personalized to you or customized to your needs, it's going to be more interesting and more reliable (...) (Male, 24, normal BMI)
- Gamification
    - 4. Whether to have one or multiple buddies, the choice depends on what works for the person. Maybe you can personalize it in some way. Maybe you can elect [to have] only one partner, or (...) to be put in a group. (Male, 20, normal BMI)
    - 5. You earn badges, which are just like token little things, and for some reason they just hook me in. (Female, 25, higher BMI)
    - 6. It loads your progress with a bar around the circle, then when it gets full it flashes. Like it's a celebration. (Female, 23, normal BMI)

Discussion

Principal Findings

In this mixed methods study, we found that participants in the overweight-obese group significantly lowered their BMI by almost 1 kg/m^2 (3%) during the first month of the intervention, showing no statistically significant difference at the end of the intervention. Participants who were overweight-obese used the app significantly less compared with individuals in the underweight-normal BMI group, mentioning negative feelings and demotivation arising from social comparison, particularly upward comparison (with fitter people). Participants in the underweight-normal BMI group were avid users of self-monitoring, feedback, and social features within the app and significantly increased their daily step count over the 6-month study duration (by more than 2000 steps). Most participants mentioned the desire for a more personalized intervention.

Social Comparison and Networking

In our study, social comparison and competition were preferred by participants in the underweight-normal BMI group (particularly upward comparison). Participants in the overweight-obese group mentioned a sense of vulnerability in sharing their data (eg, weight) in a social network and highlighted negative feelings and demotivation from upward comparison, similar to other studies [42], as well as a preference for face-to-face connections for support and accountability. These findings suggest that this group may benefit from being in a network of people with similar characteristics and goals. Previous research on weight and physical activity interventions has revealed that people seem more willing to participate in an online social network with others having common aspects (eg, real-world connections, similar goals, or fitness levels) [43-48].

This study builds on previous literature showing that online social networks can facilitate behavior change [15]. However, we found that user preferences regarding social features seem to be mixed, which indicates that such features should be available but optional, allowing users to control what information is shared and with whom [43,49,50]. Some people reported that they enjoy sharing their fitness achievements to receive praise and social support, in which case broadcasting to larger social networks such as Facebook seems to be helpful [51]. On the contrary, others did not like to share or participate in a social network but enjoyed *lurking*, that is, passively looking at content in social media without actively engaging [52]. This *digital people watching* effect promulgated by social media seemed to be enjoyed by some participants in our study.

Self-Monitoring

Regular self-monitoring of weight and BMI was an activity that most participants in our study were not particularly fond of. Participants in the overweight-obese group recognized the importance of weight monitoring but still preferred focusing on steps, due to an increased sense of perceived control over changes and the desire to avoid negative emotions. Other studies have reported on users' concerns about the potential for negative emotional and motivational impact when discrepancies between reality and goals are revealed in health apps [43]. Despite these concerns, regular self-monitoring of weight has been associated with weight loss and weight maintenance in individuals who are overweight and obese and is considered a healthy weight control strategy in people with a normal weight, with little evidence of adverse effects [53-58]. Given the importance of self-monitoring in behavior change [10,12], new strategies should be explored to promote engagement, such as decreasing the frequency of weighing to a level that individuals are able to maintain.

Personalization and Gamification

Personalization was a recurrent topic endorsed by participants in our interviews and has been emphasized in other studies of digital interventions as a critical aspect for weight management and physical activity [49,50,59]. Nowadays, technology allows for increasing levels of personalization, having the potential to enhance engagement with digital health interventions [60] by molding them to be more relevant to users based on their personal characteristics or their motivation to change lifestyle [43,61]. In our study, people within and between different BMI groups showed different preferences regarding core features such as self-monitoring and social comparison, which, in turn, seemed to influence their engagement with the intervention. In addition, there was higher variability in the rate of BMI change across the overweight-obese group compared with the underweight-normal group over the duration of the intervention, suggesting that the overweight-obese group might benefit from higher degrees of personalization to accommodate the greater variability in this group. Currently, most commercial activity trackers and mobile apps offer one-size-fits-all interventions with minimal personalization, which may be a factor in their high drop-off rates [62,63].

Strengths and Limitations

The strengths of this study include the objective measurement of outcomes using digital devices instead of self-reported data, the use of exploratory linear mixed effects analysis to better understand participants' BMI changes weekly throughout the intervention period and complement pre-post data, and the use of postintervention interviews and focus groups to better understand pilot test results and assess the advantages and disadvantages of the intervention and its components. However, the findings should be interpreted in the context of some limitations. This was a quasi-experimental study with a single-arm pre-post design, and causation cannot be inferred from our results. Posthoc subgroup analyses and linear mixed effects analysis were exploratory and might be subject to type I error. As in other studies [15,18,64], engagement with the intervention decreased over the 6-month period, and there was a high degree of missing data over the duration of the study, which affected the quality of the mixed effects analysis. Baseline daily step count (average for the first week of the study before being able to access the study app) was considerably high in both BMI groups, which might reflect the novelty effect of starting to use an activity tracker. There was a predominance of men in the higher BMI categories at baseline, and a predominance of women in the lower BMI categories, which may explain the observed higher weight loss in men. We only used step count as a measure of physical activity, and other measures (eg, minutes of moderate-to-vigorous physical activity) might have shown different results. The results of our study should be interpreted in the context of concomitant use of the Fitbit app (which provided additional features such as goal setting, not available in the fit.healthy.me app) and the different goals participants may have had for joining the study (eg, increase physical activity, lose weight).

Implications

Our study found different effects of the intervention in the underweight-normal BMI group and in the overweight-obese group, with participant perspectives also varying depending on the BMI group. The increase in step counts in the underweight-normal BMI group is promising given that any intensity of physical activity, including light intensity, is associated with a lower risk of premature mortality in a dose-response manner [5,65]. The short-term BMI decrease seen in the overweight-obese group may be due to the novelty factor of the intervention. Future research should explore whether it is possible to promote long-term physical activity and BMI changes in individuals who are overweight and obese with such an intervention and what types of interventions and features are associated with higher effectiveness (eg, interventions focusing on contextual and environmental factors, in addition to individual and social aspects; interventions offering personalization, such as the possibility to turn off social features or only allow for social comparison with similar individuals in terms of physical activity and BMI).

The importance of personalization was highlighted in this study by the heterogeneity of participant perspectives regarding intervention features. Smartphones and wireless trackers enable the collection of large volumes of personal data that can be leveraged to personalize interventions. Recent developments in artificial intelligence have led to the common use of machine learning models to optimize intervention content, timing, and delivery, based on users' preferences, habits, and other individual and contextual data [66-68]. Future research should explore the impact of personalized features to better match individual preferences, barriers, and needs to promote higher engagement and enhance the effectiveness of interventions.

Conclusions

A social networking mobile app connected to wireless tracking devices had different effects on participants in higher and lower BMI groups. Participants in the overweight-obese group showed a short-term decrease in their BMI that was not sustained after 1 month, and they used the app significantly less than participants in the underweight-normal BMI group, mentioning negative feelings with app use. Participants in the underweight-normal BMI group significantly increased their daily step count over the 6-month study duration. Most participants mentioned the desire for a more personalized intervention. Future research should explore the use of personalized features to better match individual preferences, barriers, and needs.

Acknowledgments

The authors thank Paige Newman and Annie YS Lau for help with the design and implementation of the study and Ying Wang, Holly Gehringer, and William Tong for help with preliminary data analyses.

This research was supported by the National Health and Medical Research Council grant APP1134919 (Centre for Research Excellence in Digital Health). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Authors' Contributions

The corresponding author (LL) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. LL is the guarantor. LL and EC conceptualized and designed the work; LL and HLT performed qualitative data collection and analysis; LL, JCQ, HLT, MAB, and EC contributed to quantitative data analysis and interpretation; LL wrote the first draft; LL, JCQ, HLT, MAB, and EC critically reviewed drafts for important intellectual content and approved the final version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
mHealth evidence reporting and assessment (mERA) guidelines and essential criteria.
[DOCX File , 18 KB - jmir_v22i12e19991_app1.docx]

Multimedia Appendix 2
Transparent reporting of evaluations with nonrandomized designs (TREND) statement checklist.
[DOCX File , 650 KB - jmir_v22i12e19991_app2.docx]

Multimedia Appendix 3
The consolidated criteria for reporting qualitative research (COREQ) checklist for reporting qualitative research.
[DOCX File , 19 KB - jmir_v22i12e19991_app3.docx]

Multimedia Appendix 4
Facebook post used as part of the recruitment strategy.
[DOCX File , 1084 KB - jmir_v22i12e19991_app4.docx]

Multimedia Appendix 5
Screenshots of the mobile application “fit.healthy.me”.
[DOCX File , 1390 KB - jmir_v22i12e19991_app5.docx]

Multimedia Appendix 6
Likelihood ratio tests to determine effect of predictors on BMI difference.
[DOCX File , 16 KB - jmir_v22i12e19991_app6.docx]

Multimedia Appendix 7
Linear mixed effects analysis of the weekly BMI change of each participant, using sex, weight baseline, and week as fixed effects.
[DOCX File , 16 KB - jmir_v22i12e19991_app7.docx]

References

1. GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Jul 6;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]

2. Kohl HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, Lancet Physical Activity Series Working Group. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul 21;380(9838):294-305. [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
3. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, Lancet Physical Activity Series 2 Executive Committee. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 2016 Sep 24;388(10051):1311-1324. [doi: 10.1016/S0140-6736(16)30383-X] [Medline: 27475266]
4. US Preventive Services Task Force, Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, et al. Behavioral weight loss interventions to prevent obesity-related morbidity and mortality in adults: us preventive services task force recommendation statement. J Am Med Assoc 2018 Sep 18;320(11):1163-1171. [doi: 10.1001/jama.2018.13022] [Medline: 30326502]
5. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. Br Med J 2019 Aug 21;366:l4570 [FREE Full text] [doi: 10.1136/bmj.l4570] [Medline: 31434697]
6. Guthold R, Stevens G, Riley L, Bull F. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
7. Global Mobile Consumer Survey: UK Cut. Deloitte. 2019. URL: https://www2.deloitte.com/uk/en/pages/technology-media-and-telecommunications/articles/mobile-consumer-survey.html [accessed 2020-12-01]
8. Global Mobile Consumer Survey: US Edition. Deloitte. 2018. URL: https://www2.deloitte.com/tr/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-survey-us-edition.html [accessed 2020-12-01]
9. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Mar 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
10. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
11. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
12. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
13. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
14. Thomas JG, Bond DS, Raynor HA, Papandonatos GD, Wing RR. Comparison of smartphone-based behavioral obesity treatment with gold standard group treatment and control: a randomized trial. Obesity (Silver Spring) 2019 Apr;27(4):572-580 [FREE Full text] [doi: 10.1002/oby.22410] [Medline: 30779333]
15. Laranjo L, Arguel A, Neves A, Gallagher A, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
16. Cobb NK, Graham AL. Health behavior interventions in the age of facebook. Am J Prev Med 2012 Nov;43(5):571-572. [doi: 10.1016/j.amepre.2012.08.001] [Medline: 23079184]
17. Centola D. Social media and the science of health behavior. Circulation 2013 May 28;127(21):2135-2144. [doi: 10.1161/CIRCULATIONAHA.112.101816] [Medline: 23716382]
18. Maher CA, Lewis LK, Ferrar K, Marshall S, de Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Mar 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
19. Looyestyn J, Kernot J, Boshoff K, Maher C. A web-based, social networking beginners' running intervention for adults aged 18 to 50 years delivered via a Facebook group: randomized controlled trial. J Med Internet Res 2018 Feb 26;20(2):e67 [FREE Full text] [doi: 10.2196/jmir.7862] [Medline: 29483065]
20. Tong HL, Laranjo L. The use of social features in mobile health interventions to promote physical activity: a systematic review. NPJ Digit Med 2018;1:43 [FREE Full text] [doi: 10.1038/s41746-018-0051-3] [Medline: 31304323]
21. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]
22. Ashrafian H, Toma T, Harling L, Kerr K, Athanasiou T, Darzi A. Social networking strategies that aim to reduce obesity have achieved significant although modest results. Health Aff (Millwood) 2014 Sep;33(9):1641-1647. [doi: 10.1377/hlthaff.2014.0370] [Medline: 25201670]

23. Poirier J, Bennett WL, Jerome GJ, Shah NG, Lazo M, Yeh H, et al. Effectiveness of an activity tracker- and internet-based adaptive walking program for adults: a randomized controlled trial. J Med Internet Res 2016 Mar 9;18(2):e34 [FREE Full text] [doi: 10.2196/jmir.5295] [Medline: 26860434]
24. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Effects of three motivationally targeted mobile device applications on initial physical activity and sedentary behavior change in midlife and older adults: a randomized trial. PLoS One 2016;11(6):e0156370 [FREE Full text] [doi: 10.1371/journal.pone.0156370] [Medline: 27352250]
25. Zhang J, Brackbill D, Yang S, Centola D. Efficacy and causal mechanism of an online social media intervention to increase physical activity: Results of a randomized controlled trial. Prev Med Rep 2015;2:651-657 [FREE Full text] [doi: 10.1016/j.pmedr.2015.08.005] [Medline: 26844132]
26. Hartmann-Boyce J, Johns DJ, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Effect of behavioural techniques and delivery mode on effectiveness of weight management: systematic review, meta-analysis and meta-regression. Obes Rev 2014 Jul;15(7):598-609 [FREE Full text] [doi: 10.1111/obr.12165] [Medline: 24636238]
27. Patel MS, Volpp KG, Rosin R, Bellamy SL, Small DS, Fletcher MA, et al. A randomized trial of social comparison feedback and financial incentives to increase physical activity. Am J Health Promot 2016 Jul;30(6):416-424 [FREE Full text] [doi: 10.1177/0890117116658195] [Medline: 27422252]
28. McConnon A, Kirk SF, Ransley JK. Process evaluation of an internet-based resource for weight control: use and views of an obese sample. J Nutr Educ Behav 2009;41(4):261-267. [doi: 10.1016/j.jneb.2008.07.008] [Medline: 19508931]
29. Laranjo L, Lau AY, Martin P, Tong HL, Coiera E. Use of a mobile social networking intervention for weight management: a mixed-methods study protocol. BMJ Open 2017 Jul 12;7(7):e016665 [FREE Full text] [doi: 10.1136/bmjopen-2017-016665] [Medline: 28706104]
30. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. Br Med J 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
31. des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004 Mar;94(3):361-366. [doi: 10.2105/ajph.94.3.361] [Medline: 14998794]
32. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
33. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
34. Tong HL, Coiera E, Laranjo L. Using a mobile social networking app to promote physical activity: a qualitative study of users' perspectives. J Med Internet Res 2018 Dec 21;20(12):e11439 [FREE Full text] [doi: 10.2196/11439] [Medline: 30578201]
35. Tong HL, Coiera E, Tong W, Wang Y, Quiroz JC, Martin P, et al. Efficacy of a mobile social networking intervention in promoting physical activity: quasi-experimental study. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12181 [FREE Full text] [doi: 10.2196/12181] [Medline: 30920379]
36. Yen H, Chiu H. The effectiveness of wearable technologies as physical activity interventions in weight control: A systematic review and meta-analysis of randomized controlled trials. Obes Rev 2019 Oct;20(10):1485-1493. [doi: 10.1111/obr.12909] [Medline: 31342646]
37. Biesanz JC, Deeb-Sossa N, Papadakis AA, Bollen KA, Curran PJ. The role of coding time in estimating and interpreting growth curve models. Psychol Method 2004;9(1):30-52. [doi: 10.1037/1082-989x.9.1.30]
38. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J Stat Soft 2015;67(1):1. [doi: 10.18637/jss.v067.i01]
39. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. Br Med J 2000 Jan 8;320(7227):114-116 [FREE Full text] [doi: 10.1136/bmj.320.7227.114] [Medline: 10625273]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Corcoran K, Crusius J, Mussweiler T. Social comparison: motives, standards, and mechanisms. In: Theories in Social Psychology. New York, USA: Wiley Blackwell; 2011:119-139.
42. Haferkamp N, Krämer NC. Social comparison 2.0: examining the effects of online profiles on social-networking sites. Cyberpsychol Behav Soc Netw 2011 May;14(5):309-314. [doi: 10.1089/cyber.2010.0120] [Medline: 21117976]
43. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
44. Maher C, Ryan J, Ambrosi C, Edney S. Users' experiences of wearable activity trackers: a cross-sectional study. BMC Public Health 2017 Nov 15;17(1):880 [FREE Full text] [doi: 10.1186/s12889-017-4888-1] [Medline: 29141607]

45. Cavallo DN, Tate DF, Ward DS, DeVellis RF, Thayer LM, Ammerman AS. Social support for physical activity-role of Facebook with and without structured intervention. Transl Behav Med 2014 Dec;4(4):346-354 [FREE Full text] [doi: 10.1007/s13142-014-0269-9] [Medline: 25584083]
46. Hwang KO, Ottenbacher AJ, Green AP, Cannon-Diehl MR, Richardson O, Bernstam EV, et al. Social support in an internet weight loss community. Int J Med Inform 2010 Jan;79(1):5-13 [FREE Full text] [doi: 10.1016/j.ijmedinf.2009.10.003] [Medline: 19945338]
47. Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011 Dec 2;334(6060):1269-1272 [FREE Full text] [doi: 10.1126/science.1207055] [Medline: 22144624]
48. Meng J. Your health buddies matter: preferential selection and social influence on weight management in an online health social network. Health Commun 2016 Dec;31(12):1460-1471. [doi: 10.1080/10410236.2015.1079760] [Medline: 27055008]
49. Lyzwinski LN, Caffery LJ, Bambling M, Edirippulige S. Consumer perspectives on mHealth for weight loss: a review of qualitative studies. J Telemed Telecare 2017 Feb 9;24(4):290-302. [doi: 10.1177/1357633x17692722]
50. Tang J, Abraham C, Stamp E, Greaves C. How can weight-loss app designers' best engage and support users? A qualitative investigation. Br J Health Psychol 2015 Mar;20(1):151-171. [doi: 10.1111/bjhp.12114] [Medline: 25130682]
51. Chang RC, Lu H, Yang P, Luarn P. Reciprocal reinforcement between wearable activity trackers and social network services in influencing physical activity behaviors. JMIR Mhealth Uhealth 2016 Jul 5;4(3):e84 [FREE Full text] [doi: 10.2196/mhealth.5637] [Medline: 27380798]
52. Preece J, Nonnecke B, Andrews D. The top five reasons for lurking: improving community experiences for everyone. Comput Hum Behav 2004 Mar;20(2):201-223. [doi: 10.1016/j.chb.2003.10.015]
53. Pourzanjani A, Quisel T, Foschini L. Adherent use of digital health trackers is associated with weight loss. PLoS ONE 2016 Apr 6;11(4):e0152504. [doi: 10.1371/journal.pone.0152504]
54. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017 May 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
55. Wing RR, Tate DF, Gorin AA, Raynor HA, Fava JL. A self-regulation program for maintenance of weight loss. N Engl J Med 2006 Oct 12;355(15):1563-1571. [doi: 10.1056/nejmoa061883]
56. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. Daily self-weighing and adverse psychological outcomes: a randomized controlled trial. Am J Prev Med 2014 Jan;46(1):24-29 [FREE Full text] [doi: 10.1016/j.amepre.2013.08.006] [Medline: 24355668]
57. Wing RR, Tate D, LaRose JG, Gorin AA, Erickson K, Robichaud EF, et al. Frequent self-weighing as part of a constellation of healthy weight control practices in young adults. Obesity (Silver Spring) 2015 May;23(5):943-949 [FREE Full text] [doi: 10.1002/oby.21064] [Medline: 25865175]
58. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Mar;23(2):256-265 [FREE Full text] [doi: 10.1002/oby.20946] [Medline: 25521523]
59. Broekhuizen K, Kroeze W, van Poppel MN, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286 [FREE Full text] [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
60. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
61. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The effectiveness of a web-based computer-tailored physical activity intervention using FitBit activity trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [FREE Full text] [doi: 10.2196/11321] [Medline: 30563808]
62. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
63. Ledger D, McCaffrey D. Inside Wearables: How the Science of Human Behavior Change Offers the Secret to Long-Term Engagement. New York, USA: Endeavour Partners LLC; 2014.
64. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
65. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 May 29:- [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
66. Chen JH, Asch SM. Machine learning and prediction in medicine — beyond the peak of inflated expectations. N Engl J Med 2017 Jun 29;376(26):2507-2509. [doi: 10.1056/nejmp1702071]
67. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]

68. Insel TR. Digital phenotyping: technology for a new science of behavior. J Am Med Assoc 2017 Oct 3;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]

Edited by G Eysenbach; submitted 08.05.20; peer-reviewed by J Kite, S Dick; comments to author 12.06.20; revised version received 06.08.20; accepted 11.11.20; published 08.12.20.

Please cite as:
Laranjo L, Quiroz JC, Tong HL, Arevalo Bazalar M, Coiera E
A Mobile Social Networking App for Weight Management and Physical Activity Promotion: Results From an Experimental Mixed Methods Study
J Med Internet Res 2020;22(12):e19991
URL: http://www.jmir.org/2020/12/e19991/
doi:10.2196/19991
PMID:33289670

Original Paper

Mobile Health App With Social Media to Support Self-Management for Patients With Chronic Kidney Disease: Prospective Randomized Controlled Study

Wen-Yi Li[1,2], MD; Fu-Chun Chiu[2,3], MD; Jyun-Kai Zeng[4], BSc; Yao-Wei Li[4], BSc; Su-Hua Huang[5], BSc; Hui-Chin Yeh[4,6], PhD; Bor-Wen Cheng[4], PhD; Feng-Jung Yang[1,2], MD, PhD

[1]Renal Division, Department of Internal Medicine, National Taiwan University Hospital Yun Lin Branch, Douliu, Taiwan
[2]School of Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan
[3]Cardiovascular Division, Department of Internal Medicine, National Taiwan University Hospital Yun Lin Branch, Douliu, Taiwan
[4]Department of Industrial Engineering and Management, National Yunlin University of Science and Technology, Douliu, Taiwan
[5]Department of Dietetics, National Taiwan University Hospital Yun Lin Branch, Douliu, Taiwan
[6]Department of Applied Foreign Languages, National Yunlin University of Science and Technology, Douliu, Taiwan

Corresponding Author:
Feng-Jung Yang, MD, PhD
Renal Division
Department of Internal Medicine
National Taiwan University Hospital Yun Lin Branch
No. 579, Sec. 2, Yunlin Road
Douliu, 640
Taiwan
Phone: 886 5 5323911
Email: fongrong@ntu.edu.tw

Abstract

Background: Chronic kidney disease (CKD) is a global health burden. Self-management plays a key role in improving modifiable risk factors.

Objective: The aim of this study was to evaluate the effectiveness of wearable devices, a health management platform, and social media at improving the self-management of CKD, with the goal of establishing a new self-management intervention model.

Methods: In a 90-day prospective experimental study, a total of 60 people with CKD at stages 1-4 were enrolled in the intervention group (n=30) and control group (n=30). All participants were provided with wearable devices that collected exercise-related data. All participants maintained dietary diaries using a smartphone app. All dietary and exercise information was then uploaded to a health management platform. Suggestions about diet and exercise were provided to the intervention group only, and a social media group was created to inspire the participants in the intervention group. Participants' self-efficacy and self-management questionnaire scores, Kidney Disease Quality of Life scores, body composition, and laboratory examinations before and after the intervention were compared between the intervention and control groups.

Results: A total of 49 participants completed the study (25 in the intervention group and 24 in the control group); 74% of the participants were men and the mean age was 51.22 years. There were no differences in measured baseline characteristics between the groups except for educational background. After the intervention, the intervention group showed significantly higher scores for self-efficacy (mean 171.28, SD 22.92 vs mean 142.21, SD 26.36; *P*<.001) and self-management (mean 54.16, SD 6.71 vs mean 47.58, SD 6.42; *P*=.001). Kidney Disease Quality of Life scores were also higher in the intervention group (mean 293.16, SD 34.21 vs mean 276.37, SD 32.21; *P*=.02). The number of steps per day increased in the intervention group (9768.56 in week 1 and 11,389.12 in week 12). The estimated glomerular filtration rate (eGFR) of the intervention group was higher than that of the control group (mean 72.47, SD 24.28 vs mean 59.69, SD 22.25 mL/min/1.73m^2; *P*=.03) and the decline in eGFR was significantly slower in the intervention group (–0.56 vs –4.58 mL/min/1.73m^2). There were no differences in body composition between groups postintervention.

Conclusions: The use of wearable devices, a health management platform, and social media support not only strengthened self-efficacy and self-management but also improved quality of life and a slower eGFR decline in people with CKD at stages 1-4. These results outline a new self-management model to promote healthy lifestyle behaviors for patients with CKD.

Trial Registration: ClinicalTrials.gov NCT04617431; https://www.clinicaltrials.gov/ct2/show/NCT04617431

(J Med Internet Res 2020;22(12):e19452) doi:10.2196/19452

KEYWORDS

chronic kidney disease; self-management; self-efficacy; quality of life; health management platform; wearable device

Introduction

Chronic kidney disease (CKD) is a global public health issue. CKD leads to end-stage renal disease (ESRD), and the dialysis therapy associated with ESRD incurs a huge economic burden for many countries. In Taiwan, the incidence and prevalence of ESRD are among the highest in the world [1]. Taiwan's national prevalence of CKD is estimated at 11.93% [2], which is largely due to low public awareness of the problem.

CKD is a lifelong health condition, and people with CKD often have other comorbidities such as hypertension, diabetes, and heart disease. Care for CKD involves a multidisciplinary team of researchers, engineers, and clinicians. A Cochrane database systematic review showed evidence for significant beneficial effects of regular exercise on physical fitness, walking capacity, cardiovascular dimensions (eg, blood pressure and heart rate), health-related quality of life, and nutritional parameters in adults with CKD [3]. Another systematic review showed that a combined exercise and dietary intervention resulted in a slower decline of the estimated glomerular filtration rate (eGFR) in patients with diabetes and CKD stages 3-5 [4]. However, people with CKD often feel constrained by the physical discomfort, complex treatment regimes, side effects, and liquid and dietary restrictions associated with the disease [5]. People at any stage of the disease are recommended to maintain healthy regimes of diet and exercise, along with good adherence to the medication. The ability to monitor one's lifestyle not only promotes health but also reduces the cost of health care.

Self-management is an important factor for helping people cope with chronic diseases [6]. The five fundamental self-management skills include problem-solving, decision-making, resource utilization, formation of client–health care provider partnerships, and taking action [7]. Self-efficacy is a crucial mediator between knowledge and self-care [8], and providing effective self-management support is a key policy that aims to improve the skills and confidence of patients in managing their illness [9].

In 2008, Costantini et al [10] examined the self-management experiences of people with mild to moderate CKD (stages 1-3). They found that participants wanted to self-manage their illness in collaboration with their health care providers; thus, people with early CKD need guidance and support from health care professionals to successfully self-manage their treatment [10]. In 2018, Wu et al [11] demonstrated that an innovative self-management intervention effectively decreased the serum creatinine levels and depression symptoms in people with CKD. Strengthening self-management skills has been shown to lead to a delay in the progression of CKD through improving modifiable risk factors [12].

A health management platform uses information and big data to provide the user with the ability to analyze, detect, monitor, and control risk factors of a disease. Hardinge et al [13] used a home-based mobile health (mHealth) platform for reporting daily symptoms and medication use, and for measuring physiological variables such as pulse rate and oxygen saturation in patients with chronic obstructive pulmonary disease, providing evidence for integrating telehealth interventions with clinical care pathways to support self-management.

The use of apps—software programs that run on mobile devices—is a recent approach for delivering health information and education to patients [14]. Studies have shown that a diabetes-related smartphone app combined with weekly SMS text messaging support from a health care professional could significantly improve glycemic control in adults with type 1 diabetes [15]. In a prior study, we found potential of a primary physician–led telehealth care model based on a social network service in delaying dialysis initiation for patients with stage-5 CKD [16]. According to a survey from app platforms, 67 out of 177 apps were recommended for CKD patients [17]. The most common functionalities used were CKD information and self-management (57%), e-consultation (25%), and CKD nutrition education (24%). However, the continuity of patient-centered care for CKD provided by mHealth apps is currently inadequate [10].

According to 2017 data from the Institute for Information Industry, in Taiwan, 60.2% of people older than 55 years owned a smartphone [18]. LINE is the predominant text messaging app, which is used by approximately 66.6% of people in Taiwan in their daily lives. However, there have been few interventional studies conducted to date involving the use of either wearable devices to quantify motor performance or an app-based platform to interact with CKD patients. Thus, we conducted this study to evaluate the ability of a health management platform with a wearable device and social media platform to improve participants' self-management abilities and delay the progression of CKD.

Methods

Study Design

This study was a two-arm randomized controlled trial with a pretest-posttest design. The study protocol was approved by the Research Ethics Committee of National Taiwan University Hospital (No. 201808094RINB). The trial is registered at

ClinicalTrials.gov (NCT04617431). Written informed consent was obtained from all participants before starting the study. All research procedures followed the directives of the Declaration of Helsinki.

Study Population

CKD is defined based on abnormalities of kidney function or structure for more than 3 months. The different stages of CKD form a continuum as follows [19]: stage 1, kidney damage with normal or increased eGFR (>90 mL/min/1.73 m^2); stage 2, mild reduction in eGFR (60-89 mL/min/1.73 m^2); stage 3a, moderate reduction in eGFR (45-59 mL/min/1.73 m^2); stage 3b, moderate reduction in eGFR (30-44 mL/min/1.73 m^2); stage 4, severe reduction in eGFR (15-29 mL/min/1.73 m^2); and stage 5, kidney failure (eGFR <15 mL/min/1.73 m^2 or dialysis).

This study prospectively enrolled patients with CKD stages 1-4 from the nephrology outpatient clinic of National Taiwan University Hospital Yunlin branch between January 2019 and May 2019, which is a regional teaching hospital located in a suburban area in central Taiwan. The patients were cared for by their primary care nephrologists according to the Kidney Disease Outcomes Quality Initiative guidelines [20]. Inclusion criteria were aged ≥20 years and a diagnosis of CKD stage 1-4. Those who agreed to participate in the study signed informed consent forms. Exclusion criteria were an inability to use a smartphone, impaired walking capacity, a psychiatric disorder, or any hospitalization during the previous 3 months. Participants were assigned randomly to the intervention group or the control group. We performed stratified sampling according to CKD stage and enrolled more participants at stages 2 and 3 than at stage 4. The nephrologists were blind to group allocation, whereas the investigators and participants were not.

Sample Size

G-Power 3.1.9.4 was used to calculate the sample size on the basis of an effect size of 0.35 according to the related literature [21]. To obtain a power of 0.80, α of .05, and effect size of 0.35, according to a two-tailed test, the required sample size was determined to be at least 44 (with 22 participants in each group). We estimated a 20% attrition rate, and therefore the total sample required at least 53 participants (27 in each group).

Data Collection

A brief interview was conducted with all participants, assisted by an electronic medical record search, to document their demographic profiles and comorbidities. The diagnosis of any comorbidity was documented by clinically relevant history, medical examinations, or pathological reports. Body composition was assessed using a body composition analyzer (Omron HBF-701), including body height, body weight, and percentage body fat measurements. Participants completed self-efficacy and self-management questionnaires and the Kidney Disease Quality of Life survey (KDQOL-SF), which includes the 36-item Short Form Health Survey. Laboratory data, including hemogram, serum biochemistry, electrolyte profile, and renal function assay, were measured as per routine care for CKD patients according to the guidelines of the Taiwan Society of Nephrology.

Outcome Measurement Instruments

The self-efficacy questionnaire [22] is composed of eight subscales: blood sugar or blood pressure control, diet, exercise, medication, lifestyle, infection prevention, problem-solving, and partnership. The scale has 20 questions; scores range from 0 to 10 and reflect the level of confidence, where 0 is “completely without confidence” and 10 is “completely confident.” The total scores ranged between 0 and 200, and a higher score indicates that the patient has greater confidence in controlling their disease. The Cronbach α was .81.

The self-management questionnaire [22] is composed of four subscales: partnership, compliance, self-care, and problem-solving. There are 16 questions with 4-point Likert items (1, “never”; 2, “sometimes”; 3, “usually”; and 4, “always”). The total scores ranged between 16 and 64, and a higher score indicates better self-management ability. The Cronbach α was .81.

Quality of life was measured by the KDQOL-SF [23] (dialysis version). This instrument was developed for patients with kidney disease who are on dialysis and has been adapted for nondialysis patients [24]. The KDQOL-SF includes the 36-item Short Form Health Survey, supplemented with multi-item scales targeted at particular concerns of individuals with kidney disease (eg, symptoms/problems, effects of kidney disease on daily life, burden of kidney disease, cognitive function, work status, sexual function, quality of social interaction, and sleep). The total scores ranged between 74 and 360, with a higher score indicates a better quality of life. The Cronbach α was between .61 and .90.

Intervention

Each participant was provided with a wearable device (Heart Rate Smart Wristband, GSH405-B6, Golden Smart Home Technology Corporation) (Figure 1). The wristband was approved by the National Communications Commission of Taiwan (NCC verification code: CCAB16LP1430T3). This wristband can detect steps (0-120,000 steps, division 1 step), calories, and sleep, and was validated in our previous study [25].

Figure 1. Heart Rate Smart Wristband, GSH405-B6, Golden Smart Home Technology Corporation.

Each participant downloaded the WowGoHealth app (Figure 2) to connect with the health management platform (GSH AI health platform). Participants’ exercise-related data, including the number of steps walked, distance, consumed calories, and heart rate, were collected by the wearable devices. All participants were taught how to record a dietary diary (taking photos of meals) using a smartphone app. All collected information was uploaded to the health management platform. Only the researchers could access the data on the health management platform.

Figure 2. User interface of the health management platform software (WowGoHealth app).

LINE is a mobile app operated by LINE Corporation. All users can use texts, images, videos, and audio for contact at any time. A LINE group was created to deliver medical knowledge of diet and exercise. The messages were guided by a diet manual for kidney disease (edited by the Department of Dietetics, National Taiwan University Hospital Yunlin Branch) [26].

The intervention involved diet, exercise, and self-management education. The researchers, who had been trained by a dietitian, reminded the intervention group to upload their dietary diary every day and provided suggestions about diet and exercise. In a recent systematic review and meta-analysis, covering 19 studies published from 2005 through 2017, the follow-up duration of studies related to self-management ranged from 3 months (5 studies) to 60 months [27]. This study required daily recommendations of diet and exercise; therefore, we performed the intervention for 90 days. A daily target of 7500 steps was set [28] and used to emphasize the correct concepts about exercise. The LINE app was used to inspire the participants when achieving the target number of steps. Participants in the intervention group also had the opportunity to ask questions about CKD management via the LINE app, and teleconsultations of health information were provided.

Routine care was defined as health education provided by case managers based on the national multidisciplinary pre-ESRD care project and early CKD programs [29]. The case managers offered health guidance during each outpatient visit according to the patient's renal function and blood test results. At the end of study, we provided every participant with the diet manual for kidney disease, but did not invite the control group to join the LINE group established for the intervention or provide individualized dietary suggestions.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 22.0.0 (IBM Corporation), and a two-sided *P* value <.05 was considered statistically significant. The distributional properties of data are expressed as mean (SD) for continuous variables with a normal distribution or as median (IQR) for variables with a skewed distribution. For numerical data, a Student *t* test *was used for comparisons within groups and* between groups; for categorical variables with percentages, the chi-square or Fisher exact test was used. **Analysis of covariance** (ANCOVA), a generalized linear model, was used to analyze the effectiveness of the intervention. The results were analyzed according to the per-protocol principle.

Results

Baseline Participant Characteristics

Figure 3 shows the recruitment process for this randomized controlled trial. A total of 60 participants completed the pretest, with 30 participants in each group. Five patients in the intervention group and six patients in the control group withdrew from the study. A total of 49 participants completed the posttest, including 25 in the intervention group and 24 in the control group. The baseline characteristics of the 11 patients that withdrew from the study did not differ from those of the remaining 49 participants.

The mean age of all participants was 51.22 years (SD 10.98) and 73.5% were men. The baseline characteristics of these participants are presented in Table 1. The patients with stage-2 CKD outnumbered those with stages 3 and 4. There were no differences in comorbidities between groups. The patients in the intervention and control groups had similar levels of serum creatinine. Most of the participants had a bachelor's degree, although more participants in the intervention group had a university degree compared with the control group.

Figure 3. Flow diagram of participant recruitment and randomization.

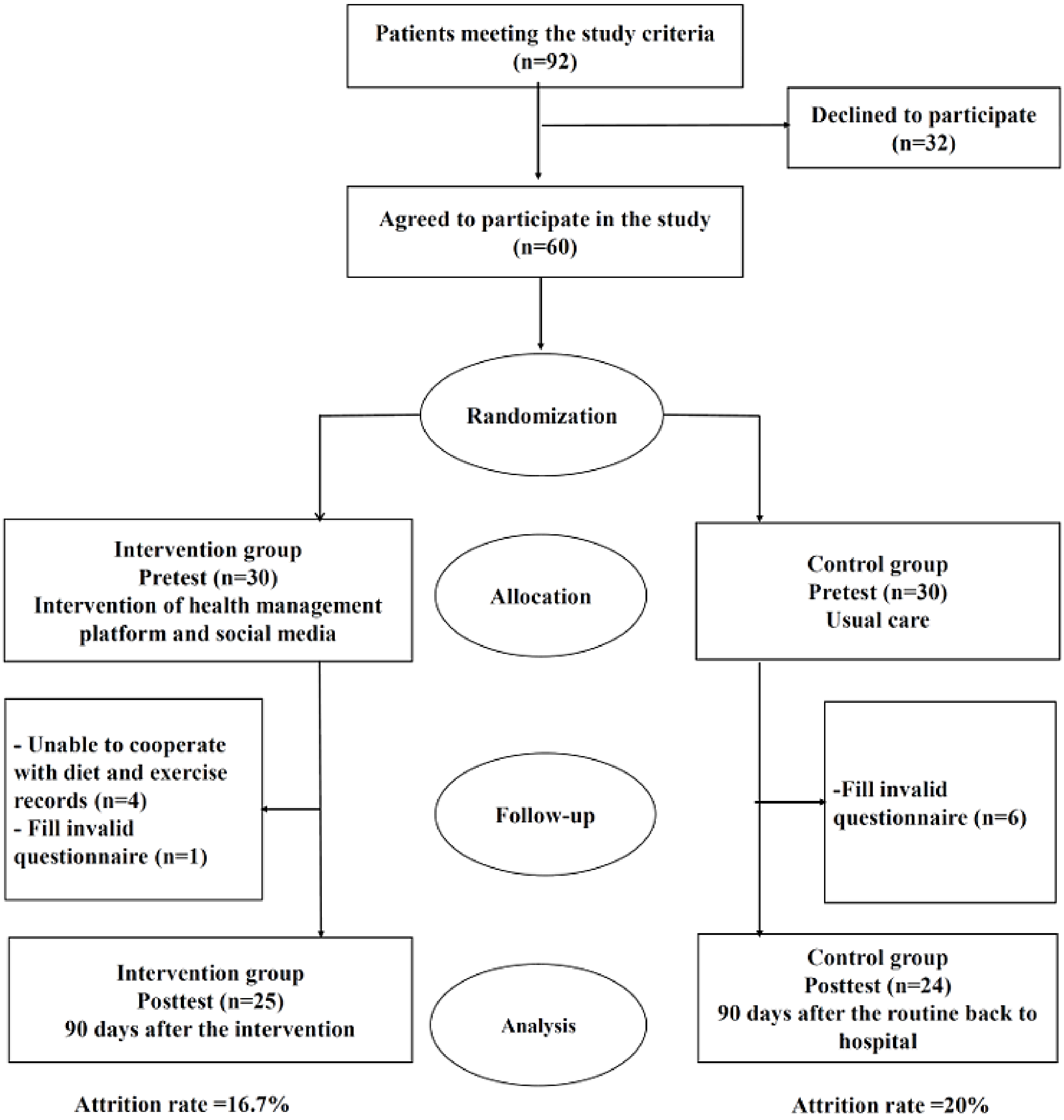


Table 1. Baseline characteristics of the intervention and control groups (N=49).

Characteristic	All (N=49)	Intervention group (n=25)	Control group (n=24)	*P* value
Men, n (%)	36 (74)	17 (68)	19 (79)	.52
Age (years), mean (SD)	51.22 (10.98)	50.60 (11.87)	51.87 (10.20)	.69
Aged ≥65 years, n (%)	4 (8)	3 (12)	1 (4)	.48
Education level, n (%)				.02
Elementary school	2 (4)	0 (0)	2 (8)	
Junior high school	5 (10)	1 (4)	4 (17)	
Senior high school	13 (27)	4 (16)	9 (38)	
College or university	29 (59)	20 (80)	9 (38)	
Comorbidities, n (%)				
Diabetes mellitus	17 (35)	8 (32)	9 (38)	.50
Hypertension	23 (47)	11 (44)	12 (50)	.39
Dyslipidemia	32 (65)	16 (64)	16 (67)	.50
Body composition, mean (SD)				
Body weight (kg)	76.27 (14.29)	75.84 (15.52)	76.71 (13.19)	.84
Body fat percentage	29.39 (5.51)	29.90 (5.93)	28.84 (5.11)	.51
BMI	27.28 (4.29)	27.07 (4.46)	27.05 (4.19)	.73
Basal metabolic rate	1638.29 (246.50)	1619.6 (271.51)	1657.75 (221.61)	.59
Laboratory parameters, mean (SD)				
Creatinine (mg/dL)	1.26 (0.40)	1.15 (0.34)	1.33 (0.44)	.12
eGFR[a] (mL/min/1.73 m^2)	66.53 (23.61)	73.03 (25.01)	64.27 (22.72)	.21
CKD[b] stage, n (%)				.46
1	8 (16)	5 (20)	3 (13)	
2	22 (45)	11 (44)	11 (46)	
3a	11 (22)	7 (28)	4 (17)	
3b	7 (14)	2 (8)	5 (21)	
4	1 (2)	0 (0)	1 (4)	

[a] eGFR: estimated glomerular filtration rate.

[b] CKD: chronic kidney disease.

Body Composition and Exercise

As shown in Table 2, there was no difference in baseline body weight in the control and intervention groups (*P*=.84). The baseline BMI and body fat percentage levels were similar between the two groups (*P*=.73 and *P*=.51, respectively). At the end of the study, both groups showed modest weight gains with no differences between groups (*P*=.89). There were no differences in body composition (body fat percentage, basal metabolic rate) between the two groups at the end of the study.

Table 2. Comparison of body composition between groups.

Variable	Intervention group (n=25)		Control group (n=24)		ANCOVA[a]		Partial ε^2	R^2 (adjusted R^2)	Power
	Pretest, mean (SD)	Posttest, mean (SD)	Pretest, mean (SD)	Posttest, mean (SD)	*F* value (df=1)	*P* value			
Body weight (kg)	75.84 (15.52)	76.25 (15.52)	76.71 (13.19)	76.80 (13.27)	0.294	.59	0.006	0.982 (0.981)	0.083
Body fat percentage	29.90 (5.93)	29.44 (6.18)	28.84 (5.11)	29.02 (5.87)	0.970	.33	0.021	0.858 (0.852)	0.161
BMI (kg/m^2)	27.07 (4.46)	27.19 (4.45)	27.05 (4.19)	27.73 (3.91)	0.181	.67	0.004	0.918 (0.914)	0.070
Basal metabolic rate (kJ/m^2·h)	1619.60 (271.51)	1620.08 (280.31)	1657.75 (221.61)	1648.12 (232.62)	0.325	.57	0.007	0.941 (0.938)	0.086

[a] ANCOVA: analysis of covariance.

After the 90-day intervention, the steps per day increased in the intervention group (9768.56 in the 1st week and 11,389.12 in the 12th week). Although this difference was not significant (*P*=.10), it indicates a trend of increased physical activity in the intervention group but not in the control group.

Physiological Indicators After the Intervention

We used ANCOVA to analyze the differences in serum creatinine and eGFR levels between the intervention and control groups after the intervention. As shown in Table 3, serum creatinine levels were lower in the intervention group than in the control group, although the difference was not significant. The eGFR of the intervention group was significantly higher than that of the control group. In other words, the decline in eGFR was significantly slower in the intervention group (–0.56 vs –4.58 mL/min/1.73 m^2). There were no significant differences between groups in blood glucose levels or uric acid and lipid profiles.

Table 3. Comparison of physiological indicators between groups.

Variable	Intervention group (n=25)		Control group (n=24)		ANCOVA[a]		Partial ε^2	R^2 (adjusted R^2)	Power
	Pretest, mean (SD)	Posttest, mean (SD)	Pretest, mean (SD)	Posttest, mean (SD)	*F* value (df=1)	*P* value			
Creatinine (mg/dL)	1.15 (0.34)	1.16 (0.39)	1.33 (0.44)	1.51 (0.81)	1.300	.26	0.027	0.810 (0.801)	0.200
eGFR[b] (mL/min/1.73 m^2)	73.03 (25.01)	72.47 (24.28)	64.27 (22.72)	56.69 (22.25)	5.341	.03	0.104	0.921 (0.918)	0.619
Uric acid (mg/dL)	6.28 (1.29)	6.28 (1.14)	6.30 (1.63)	6.70 (1.60)	2.271	.14	0.047	0.575 (0.556)	0.314
T-CHO[c] (mg/dL)	181.80 (37.57)	179.40 (35.20)	174.42 (43.83)	169.21 (39.54)	0.553	.46	0.012	0.685 (0.672)	0.113
TG[d] (mg/dL)	170.28 (77.99)	160.20 (85.29)	257.37 (273.39)	272.87 (380.67)	0.225	.64	0.005	0.540 (0.520)	0.075
LDL-C[e] (mg/dL)	108.24 (29.78)	105.84 (27.86)	95.62 (38.25)	92.29 (31.90)	0.932	.34	0.020	0.840 (0.833)	0.157
GluAC[f] (mg/dL)	111.04 (39.12)	112.76 (41.91)	118.25 (39.13)	116.25 (28.74)	0.007	.93	0.000	0.396 (0.370)	0.051

[a] ANCOVA: analysis of covariance.

[b] eGFR: estimated glomerular filtration rate.

[c] T-CHO: total cholesterol.

[d] TG: triglyceride.

[e] LDL-C: low-density lipoprotein cholesterol.

[f] GluAC: glucose ante cibum.

Self-Efficacy and Self-Management

Before the intervention, there was no significant difference in the mean self-efficacy scores between the intervention and control groups; however, after the 90-day intervention, self-efficacy scores were significantly higher in the intervention group (Table 4). Among the self-efficacy subscales, only the infection prevention subscale showed no significant difference between groups. In the other 7 subscales, the intervention group showed significant improvement, especially in blood sugar or blood pressure control, partnership, and lifestyle (Table 4).

The baseline self-management scores also showed no significant difference between the intervention and control groups. However, after the 90-day intervention, self-management scores were significantly higher in the intervention group, and this difference was evident for all 4 subscales (Table 4).

Table 4. Comparison of self-efficacy and self-management scores between groups.

Variable	Intervention group (n=25)		Control group (n=24)		ANCOVA[a]		Partial ε^2	R^2 (adjusted R^2)	Power
	Pretest, mean (SD)	Posttest, mean (SD)	Pretest, mean (SD)	Posttest, mean (SD)	*F* value (df=1)	*P* value			
Self-efficacy scores									
Total self-efficacy of CKD[b]	167.36 (30.89)	171.28 (22.92)	154.29(29.04)	142.21 (26.36)	13.728	.001	0.230	0.388 (0.361)	0.952
Blood sugar or blood pressure control	24.36 (6.24)	26.32 (4.49)	20.58 (6.19)	19.71 (5.49)	15.393	<0.001	0.251	0.534 (0.514)	0.970
Diet	32.40 (7.37)	31.08 (7.15)	28.96 (7.42)	25.62 (6.14)	6.358	.02	0.121	0.175 (0.140)	0.695
Exercise	24.08 (6.05)	24.28 (5.25)	20.33 (7.91)	19.29 (6.28)	5.555	.02	0.108	0.308 (0.278)	0.636
Medical treatment	27.48 (4.11)	28.12 (3.00)	27.29 (4.14)	25.58 (4.22)	5.975	.02	0.115	0.169 (0.133)	0.668
Lifestyle	17.60 (3.37)	18.04 (2.21)	16.37 (3.63)	14.92 (4.61)	8.865	.005	0.162	0.569 (0.550)	0.830
Infection prevention	7.12 (3.24)	7.80 (1.91)	7.00 (2.13)	6.96 (2.05)	2.206	.14	0.046	0.119 (0.081)	0.307
Problem-solving	16.52 (4.71)	17.08 (3.01)	16.50 (3.13)	14.46 (3.69)	7.959	.007	0.147	0.213 (0.179)	0.789
Partnership	17.80 (4.09)	18.56 (2.20)	17.62 (2.90)	15.67 (3.51)	11.872	.001	0.205	0.220 (0.187)	0.921
Self-management scores									
Total self-management of CKD	52.88 (8.25)	54.16 (6.71)	49.20 (7.04)	47.58 (6.42)	8.929	.004	0.163	0.477 (0.454)	0.833
Partnership	10.52 (1.85)	10.48 (1.58)	9.79 (1.67)	8.96 (1.71)	7.894	.007	0.146	0.308 (0.278)	0.785
Compliance	6.80 (1.29)	6.88 (0.93)	6.29 (1.04)	6.17 (0.87)	5.182	.03	0.101	0.298 (0.267)	0.606
Self-care	28.96 (4.70)	30.16 (3.80)	27.00 (3.97)	27.08 (3.82)	5.170	.03	0.101	0.480 (0.458)	0.605
Problem-solving	6.60 (1.32)	6.64 (1.29)	6.12 (1.51)	5.37 (1.34)	9.589	.003	0.172	0.355 (0.327)	0.858

[a] ANCOVA: analysis of covariance.

[b] CKD: chronic kidney disease.

Quality of Life

Before the intervention, the KDQOL-SF scores did not differ significantly between the two groups (*P*=.64); however, after the 90-day intervention, the KDQOL-SF scores were significantly higher in the intervention group (Table 5). The scores for the subscales of physical functioning and medical staff encouragement were significantly higher in the intervention group, with no differences found for the other subscales (Table 5).

Table 5. Comparison of self-efficacy and self-management scores between groups.

Variable	Intervention group (n=25)		Control group (n=24)		ANCOVA[a]		Partial ε^2	R^2 (adjusted R^2)	Power
	Pretest, mean (SD)	Posttest, mean (SD)	Pretest, mean (SD)	Posttest, mean (SD)	*F* value (df=1)	*P* value			
All quality of life scales	288.92 (26.19)	293.16 (34.21)	285.04 (31.61)	276.37 (32.21)	5.716	.02	0.111	0.695 (0.681)	0.648
36-item health survey scales	114.68 (13.35)	118.00 (15.64)	108.66 (14.28)	107.63 (15.93)	2.813	.10	0.058	0.574 (0.555)	0.375
General health perceptions	20.12 (3.92)	19.52 (5.12)	18.08 (3.42)	17.75 (3.48)	0.495	.49	0.011	0.175 (0.139)	0.106
Physical functioning	25.88 (3.82)	27.24 (3.18)	24.17 (4.91)	24.04 (4.67)	6.279	.02	0.120	0.549 (0.530)	0.689
Role-physical	7.04 (1.15)	7.40 (1.29)	6.62 (1.69)	6.50 (1.77)	3.807	.06	0.076	0.543 (0.524)	0.480
Role-emotional	5.36 (1.15)	5.60 (0.96)	4.92 (1.31)	5.33 (1.09)	0.031	.86	0.001	0.383 (0.356)	0.053
Social function	8.12 (1.61)	8.44 (1.63)	8.08 (1.66)	8.08 (1.69)	0.651	.42	0.014	0.247 (0.214)	0.124
Pain	9.44 (1.73)	9.40 (1.98)	8.92 (1.95)	8.75 (2.38)	0.234	.63	0.005	0.483 (0.461)	0.076
Emotional well-being	22.00 (2.78)	23.00 (3.38)	22.25 (2.66)	21.75 (3.35)	2.115	.15	0.044	0.127 (0.089)	0.296
Energy/fatigue	16.72 (3.03)	17.40 (3.40)	15.62 (2.96)	15.42 (3.62)	2.182	.15	0.045	0.375 (0.348)	0.304
All kidney disease– targeted scales	174.24 (15.04)	175.16 (19.73)	170.12 (16.81)	168.75 (17.94)	0.600	.44	0.013	0.712 (0.699)	0.118
Burden of kidney disease	14.36 (3.69)	15.36 (3.68)	15.33 (3.96)	15.17 (3.42)	0.912	.35	0.019	0.401(0.375)	0.155
Quality of social interaction	15.48 (2.06)	14.96 (2.65)	13.62 (4.02)	14.58 (2.50)	0.004	.95	0.000	0.083 (0.044)	0.050
Cognitive function	15.28 (1.99)	15.16 (2.28)	14.71 (2.97)	14.79 (2.19)	0.039	.85	0.001	0.263 (0.231)	0.054
Symptom/problems	48.20 (5.31)	48.40 (5.96)	47.67 (5.05)	46.29 (6.35)	1.726	.20	0.036	0.509 (0.488)	0.251
Effects of kidney disease	36.12 (4.30)	35.72 (5.65)	34.83 (5.85)	35.21 (5.76)	0.264	.61	0.006	0.564 (0.545)	0.080
Sleep	19.48 (2.86)	20.28 (3.36)	19.54 (3.20)	18.79 (4.28)	2.632	.11	0.054	0.305 (0.275)	0.355
Social support	6.60 (1.47)	6.24 (1.69)	6.62 (1.34)	6.45 (0.88)	0.342	.56	0.007	0.179 (0.144)	0.088
Work status	3.64 (0.57)	3.60 (0.65)	3.42 (0.65)	3.37 (0.57)	0.393	.53	0.008	0.387 (0.361)	0.094
Patient satisfaction	6.68 (1.41)	6.48 (1.71)	6.01 (1.32)	6.08 (1.14)	0.087	.77	0.002	0.351 (0.323)	0.060
Medical staff encouragement	8.40 (1.19)	8.96 (1.02)	8.17 (1.31)	8.00 (1.38)	8.263	.006	0.152	0.418 (0.393)	0.803

[a] ANCOVA: analysis of covariance.

Dietary Suggestions and Participant Satisfaction

The investigators provided individualized dietary suggestions online for intervention group participants. The most frequent suggestions included increasing the amount and types of vegetables and fruits, reducing protein intake, and adjusting calories. Most of the participants in the intervention group (19/25, 76%) gave positive feedback about the intervention, and 88% (22/25) reported better changes in dietary and exercise habits. Six of the 25 (24%) participants in the intervention group felt that wearing the smart wristband was inconvenient.

Discussion

Principal Findings

This study provides evidence that the use of wearable devices with a health management platform and social media support not only strengthened self-efficacy and self-management but also improved quality of life and slowed the eGFR decline in patients with CKD stages 1-4. These results establish a new self-management model for promoting healthy lifestyle behaviors in patients with CKD. Although Taiwan has implemented preventive programs and patient education for people with CKD since 2006, these programs do not emphasize self-management and are not individualized.

Self-Management in CKD

Long-term CKD management requires a high level of patient involvement, both in terms of decision-making and in the implementation of care. Patient self-management approaches generally aim to increase self-sufficiency and reduce health care costs.

In a 2019 systematic review and meta-analysis, 19 studies published from 2005 through 2017 were identified, with a total of 2540 CKD patients and a mean follow up of 13.44 months [27]. Compared with usual care, self-management interventions did not show a significant difference in the risk of all-cause mortality or change in eGFR. However, self-management interventions were associated with a lower 24-hour urinary protein excretion level, lower blood pressure level, lower C-reactive protein level, and longer distance on the 6-minute walk when compared with those of controls. These four factors are all known risk factors for cardiovascular disease. Among the 19 studies, the most common intervention type was face-to-face intervention; only 5 studies used a combination of face-to-face and telehealth approaches [30-32]. No investigation in the review used a combination of wearable devices, a health management platform, and social media for an intervention without any face-to-face education, which was adopted as the intervention in our study.

We observed a slower deterioration of renal function following the intervention, which was not consistent with the results of the systematic review. However, several studies within the review showed positive effects of self-management on renal function. In a randomized controlled trial conducted in 2011, Chen et al [32] addressed self-management support in CKD stage 3-5 (mean age of participants 68.2, SD 12.1 years) with an intervention that comprised health information, patient education, weekly telephone-based support, and a support group. After 12 months, the absolute eGFR was significantly higher in the intervention group than that in the control group (29.11, SD 20.61 vs 15.72, SD 10.67 mL/min/1.73 m^2; *P*=.04), and fewer hospitalization events were noted in participants who received the intervention. In 2018, Wu et al [11] tested an innovative self-management intervention in patients with CKD stages 3b-5 (mean age 70.2, SD 11.6 years). The intervention included a video, a group training manual about self-efficacy and management of CKD, telephone interviews, and small-group interventions (once per week for 1 month). After 3 months, the intervention effectively decreased serum creatinine levels (2.96, SD 2.14 vs 3.04, SD 2.17 mg/dL; *P*=.02) and levels of depression (*P*=.02) in CKD patients [11]. The findings of our study are consistent with those of the two studies summarized above, possibly because of the similarities in the interventions, which used convenient technology and nearly daily contact between researchers and participants.

Our intervention group had better scores in self-efficacy of blood sugar and blood pressure control, which reflects a belief in their ability to take action for their chronic illness. Previous investigations have also shown improvement of health-related quality of life scores when tailored information was provided [33]. This may be related to the sense of feeling supported and empowered [10]. However, the higher self-efficacy scores in this study were not reflected in actual blood glucose levels, possibly because the study period of 90 days was too short.

Design of the Telehealth Intervention

This interventional study is unique in the combination of a wearable device, a health management platform, and social media in a population with CKD stages 1-4. Few studies have adopted wearable devices, a health management platform, and social media together to quantify motor performance with immediate feedback to empower participants [17]. The use of new technology offers another way to deliver health care. However, it requires technological literacy among the patient population.

Although there is a high prevalence of smartphone use in Taiwan, this study was performed in a rural area, which likely reduces motivation for smartphone use compared to use in an urban area. In addition, the average age of patients with CKD stages 1-4 in the hospital from which the participants were recruited is 69.72 years, whereas the participants in this study were predominantly younger (mean age 51.22 years); they were also open-minded, eager to learn, and aggressive about keeping healthy.

Our intervention type was a multifactorial behavior modification, including exercise and diet. A previous study showed that a 4-month dietary calorie restriction and aerobic exercise intervention resulted in benefits in body weight, fat mass, and markers of oxidative stress and inflammatory response in patients with moderate to severe CKD [34]. Another study found that a dietitian-provided telehealth-delivered regular dietary intervention was well accepted by patients with CKD stages 3-4 [35]; they were more aware of their dietary needs and could prioritize dietary behavior changes. In our study, dietary calorie restriction was not suggested, and body weight change was insignificant. However, we did provide real-time suggestions to the participants according to uploaded images of their current diet, and they could adjust their dietary choices day by day.

The 2015 Taiwan Chronic Kidney Disease Clinical Guidelines [36] encourage physical activity compatible with cardiovascular health and tolerance, aiming for at least 30 minutes of exercise 5 times per week. For CKD patients, there is no limitation on the type of exercise, although strenuous exercise is not recommended. A goal of 10,000 steps per day is commonly believed to be necessary for maintaining good health, but the evidence for this number is lacking. In a recent prospective cohort study of 18,289 women in the United States (aged 72.0, SD 5.7 years), as few as approximately 4400 steps per day was significantly related to lower mortality rates as compared with approximately 2700 steps per day. With more steps per day, mortality rates progressively decreased before leveling off at approximately 7500 steps per day [28]. For this reason, this study set a goal of 7500 steps per day for the participating CKD patients. In our pedometer-guided exercise study, the steps per day tracked by participants were better than our expectations.

During a global pandemic such as COVID-19, telemedicine is a good method of clinical practice to limit travel and exposure, and encourage social distancing while inspiring care from medical staff. For example, chatbots are artificial intelligence (AI) computer programs that simulate text-messaging dialogs with users. The purpose is to help users solve trivial problems in their daily lives, such as ordering food and calling a car. In the near future, AI health advice robots will be widely used to integrate services such as circuit training and diet programs with healthy food, long-term tracking of health indices, personalized AI advice, and social networking. More investigation is needed into the potential effects of AI on CKD management. Regardless of the stage of the disease, maintaining healthy regimes of diet and exercise, and good adherence to medication not only promotes health but also reduces the cost of health care.

Limitations

Participants in this study were recruited from one teaching hospital in central Taiwan. In addition, the sample was relatively small because of the limited number of wearable devices available. At baseline, we found a significant difference in educational level between the intervention and control groups. Education might thus be a major confounder for the results obtained. In an 11-year follow-up study of the Dutch general population, a low educational level was associated with an elevated risk of CKD [37]. This association is suggested to be driven by higher rates of diabetes and modifiable risk factors such as abdominal obesity, smoking, low potassium intake, and hypertension in those with lower education. In a US study (with a median follow-up time of approximately 23 years), socioeconomic status (annual household income, educational attainment, or neighborhood deprivation) was associated not only with ESRD risk but also with eGFR decline, although the association with CKD appeared to be weaker [38]. This suggests that observing the effects of educational level and socioeconomic status takes time. In this study, given that the follow-up period was only 3 months, the effect of the difference in educational level is considered to be limited.

Although self-efficacy, self-management, and quality of life outcomes improved after the 90-day intervention, the long-term effect of the intervention should also be further evaluated. The effectiveness of an intervention on body composition and renal function cannot be well evaluated over a short period of time. A follow-up period of at least 1 year would allow for better evaluation of the long-term effectiveness of the intervention.

Finally, our study and intervention design made it impossible to distinguish between the effects of exercise, dietary intervention, or emotional support via social media.

Conclusion

A self-management intervention that combines wearable devices, a health management platform, and social media could strengthen self-efficacy and self-management, and lead to improvements in quality of life for people with CKD stages 1-4. The effects of this nonpharmacologic intervention were also reflected in a slower decline in eGFR. These results outline a self-management model that can promote healthy lifestyle behaviors in patients with CKD.

Acknowledgments

This study was supported by grants from the Ministry of Science and Technology (MOST 107-2635-E-002-001-) and the National Taiwan University Hospital Yunlin Branch (NTUHYL108.F005). The authors are grateful to the Department of Medical Education and Research, National Taiwan University Hospital Yunlin Branch, for assistance with the statistical analysis.

Authors' Contributions

FY, WL, CC, and JZ designed the study; WL, JZ, and YL recruited study participants; JZ performed the statistical analyses; SH provided expert opinions for the nutritional intervention; WL, CC, FY, and CH wrote the manuscript; FY, HY, and BC provided expert opinions for the study design and edited the manuscript. All authors participated in discussion, interpretation, and final preparation of the manuscript.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was retrospectively registered in June 2020. The authors explained that because the intervention was noninvasive, the investigators were unaware of the requirement for registration. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

References

1. Saran R, Robinson B, Abbott KC, Agodoa LYC, Bragg-Gresham J, Balkrishnan R, et al. US Renal Data System 2018 Annual Data Report: Epidemiology of Kidney Disease in the United States. Am J Kidney Dis 2019 Mar;73(3 Suppl 1):A7-A8 [FREE Full text] [doi: 10.1053/j.ajkd.2019.01.001] [Medline: 30798791]
2. Wen C, Cheng TYD, Tsai MK, Chang YC, Chan HT, Tsai SP, et al. All-cause mortality attributable to chronic kidney disease: a prospective cohort study based on 462 293 adults in Taiwan. Lancet 2008 Jun 28;371(9631):2173-2182. [doi: 10.1016/S0140-6736(08)60952-6] [Medline: 18586172]
3. Heiwe S, Jacobson SH. Exercise training for adults with chronic kidney disease. Cochrane Database Syst Rev 2011 Oct 05(10):CD003236. [doi: 10.1002/14651858.CD003236.pub2] [Medline: 21975737]
4. Van Huffel L, Tomson CRV, Ruige J, Nistor I, Van Biesen W, Bolignano D. Dietary restriction and exercise for diabetic patients with chronic kidney disease: a systematic review. PLoS One 2014;9(11):e113667 [FREE Full text] [doi: 10.1371/journal.pone.0113667] [Medline: 25423489]
5. Tong A, Sainsbury P, Chadban S, Walker RG, Harris DC, Carter SM, et al. Patients' experiences and perspectives of living with CKD. Am J Kidney Dis 2009 Apr;53(4):689-700. [doi: 10.1053/j.ajkd.2008.10.050] [Medline: 19216015]
6. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
7. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
8. Wu SV, Hsieh N, Lin L, Tsai J. Prediction of self-care behaviour on the basis of knowledge about chronic kidney disease using self-efficacy as a mediator. J Clin Nurs 2016 Sep;25(17-18):2609-2618. [doi: 10.1111/jocn.13305] [Medline: 27364760]
9. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
10. Costantini L, Beanlands H, McCay E, Cattran D, Hladunewich M, Francis D. The self-management experience of people with mild to moderate chronic kidney disease. Nephrol Nurs J 2008;35(2):147-155; quiz 156. [Medline: 18472683]
11. Wu SFV, Lee MC, Hsieh NC, Lu KC, Tseng HL, Lin LJ. Effectiveness of an innovative self-management intervention on the physiology, psychology, and management of patients with pre-end-stage renal disease in Taiwan: A randomized, controlled trial. Jpn J Nurs Sci 2018 Oct;15(4):272-284. [doi: 10.1111/jjns.12198] [Medline: 29266792]
12. Walker R, Marshall M, Polaschek N. Improving self-management in chronic kidney disease: a pilot study. Renal Soc Australas J 2013;9(3):116-125 [FREE Full text]
13. Hardinge M, Rutter H, Velardo C, Shah SA, Williams V, Tarassenko L, et al. Using a mobile health application to support self-management in chronic obstructive pulmonary disease: a six-month cohort study. BMC Med Inform Decis Mak 2015 Jun 18;15:46 [FREE Full text] [doi: 10.1186/s12911-015-0171-5] [Medline: 26084626]
14. Information Resources Management Association, editor . Healthcare Policy and Reform: Concepts, Methodologies, Tools, and Applications. Hershey, PA: IGI Global; 2018.
15. Kirwan ML, Megonigal JP. Tidal wetland stability in the face of human impacts and sea-level rise. Nature 2013 Dec 05;504(7478):53-60. [doi: 10.1038/nature12856] [Medline: 24305148]
16. Yang F, Hou Y, Chang R. The Impact of a Social Networking Service-Enhanced Smart Care Model on Stage 5 Chronic Kidney Disease: Quasi-Experimental Study. J Med Internet Res 2020 Apr 14;22(4):e15565 [FREE Full text] [doi: 10.2196/15565] [Medline: 32200348]
17. Lee Y, Cui Y, Tu M, Chen Y, Chang P. Mobile Health to Maintain Continuity of Patient-Centered Care for Chronic Kidney Disease: Content Analysis of Apps. JMIR Mhealth Uhealth 2018 Apr 20;6(4):e10173 [FREE Full text] [doi: 10.2196/10173] [Medline: 29678805]
18. Spinelli G, Micocci M, Martin W, Wang Y. From medical devices to everyday products: exploring cross-cultural perceptions of assistive technology. Design for Health 2019 Nov 05;3(2):324-340. [doi: 10.1080/24735132.2019.1680065]
19. Kidney Disease: Improving Global Outcomes (KDIGO) Work Group. Chapter 1: Definition and classification of CKD. Kidney Int Suppl (2011) 2013 Jan;3(1):19-62 [FREE Full text] [doi: 10.1038/kisup.2012.64] [Medline: 25018975]
20. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney International Supplement. 2013. URL: https://kdigo.org/wp-content/uploads/2017/02/KDIGO_2012_CKD_GL.pdf [accessed 2020-12-06]
21. Lii Y, Tsay S, Wang T. Group intervention to improve quality of life in haemodialysis patients. J Clin Nurs 2007 Nov;16(11C):268-275. [doi: 10.1111/j.1365-2702.2007.01963.x] [Medline: 17931320]
22. Lee CL. Self-efficacy, self-management, and the quality of life for patients with chronic kidney disease. Master's Thesis of National Taiwan Normal University.: National Taiwan Normal University; 2013. URL: https://www.airitilibrary.com/Publication/alDetailedMesh?docid=U0021-0801201418035228#Altmetrics [accessed 2019-12-20]
23. Hays RD, Kallich JD, Mapes DL, Coons SJ, Carter WB. Development of the kidney disease quality of life (KDQOL) instrument. Qual Life Res 1994 Oct;3(5):329-338. [doi: 10.1007/BF00451725] [Medline: 7841967]
24. Cook M, Girman CJ, Stein PP, Alexander CM. Initial monotherapy with either metformin or sulphonylureas often fails to achieve or maintain current glycaemic goals in patients with Type 2 diabetes in UK primary care. Diabet Med 2007 Apr;24(4):350-358. [doi: 10.1111/j.1464-5491.2007.02078.x] [Medline: 17335466]

25. Zeng, Jyun-Kai. Application of Health Management Platform with Wearable Devices to Strengthen Self-management of Chronic Kidney Disease Patient. Master's Thesis of National Yunlin University of Science and Technology. Taiwan; 2018. URL: https://hdl.handle.net/11296/q22csw [accessed 2019-12-20]
26. Chronic Kidney Disease. Taiwan: Health Promotion Administration, Ministry of Health and Welfare URL: https://www.hpa.gov.tw/File/Attach/6639/File_6234.pdf [accessed 2020-12-06]
27. Peng S, He J, Huang J, Lun L, Zeng J, Zeng S, et al. Self-management interventions for chronic kidney disease: a systematic review and meta-analysis. BMC Nephrol 2019 Apr 26;20(1):142 [FREE Full text] [doi: 10.1186/s12882-019-1309-y] [Medline: 31027481]
28. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of Step Volume and Intensity With All-Cause Mortality in Older Women. JAMA Intern Med 2019 May 29;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
29. Wu MY, Wu MS. Taiwan renal care system: A learning health-care system. Nephrology (Carlton) 2018 Oct;23(Suppl 4):112-115. [doi: 10.1111/nep.13460] [Medline: 30298659]
30. Campbell KL, Ash S, Davies PS, Bauer JD. Randomized controlled trial of nutritional counseling on body composition and dietary intake in severe CKD. Am J Kidney Dis 2008 May;51(5):748-758. [doi: 10.1053/j.ajkd.2007.12.015] [Medline: 18436085]
31. Ishani A, Christopher J, Palmer D, Otterness S, Clothier B, Nugent S, Center for Innovative Kidney Care. Telehealth by an Interprofessional Team in Patients With CKD: A Randomized Controlled Trial. Am J Kidney Dis 2016 Jul;68(1):41-49. [doi: 10.1053/j.ajkd.2016.01.018] [Medline: 26947216]
32. Chen S, Tsai Y, Sun C, Wu I, Lee C, Wu M. The impact of self-management support on the progression of chronic kidney disease--a prospective randomized controlled trial. Nephrol Dial Transplant 2011 Nov;26(11):3560-3566. [doi: 10.1093/ndt/gfr047] [Medline: 21414969]
33. Blakeman T, Blickem C, Kennedy A, Reeves D, Bower P, Gaffney H, et al. Effect of information and telephone-guided access to community support for people with chronic kidney disease: randomised controlled trial. PLoS One 2014;9(10):e109135 [FREE Full text] [doi: 10.1371/journal.pone.0109135] [Medline: 25330169]
34. Ikizler TA, Robinson-Cohen C, Ellis C, Headley SA, Tuttle K, Wood RJ, et al. Metabolic Effects of Diet and Exercise in Patients with Moderate to Severe CKD: A Randomized Clinical Trial. J Am Soc Nephrol 2018 Jan;29(1):250-259 [FREE Full text] [doi: 10.1681/ASN.2017010020] [Medline: 29038285]
35. Warner MM, Tong A, Campbell KL, Kelly JT. Patients' Experiences and Perspectives of Telehealth Coaching with a Dietitian to Improve Diet Quality in Chronic Kidney Disease: A Qualitative Interview Study. J Acad Nutr Diet 2019 Aug;119(8):1362-1374. [doi: 10.1016/j.jand.2019.01.023] [Medline: 30979633]
36. Taiwan Chronic Kidney Disease Clinical Guidelines. Miaoli County, Taiwan: National Health Research Institutes; 2015.
37. Thio C, Vart P, Kieneker LM, Snieder H, Gansevoort RT, Bültmann U. Educational level and risk of chronic kidney disease: longitudinal data from the PREVEND study. Nephrol Dial Transplant 2020 Jul 01;35(7):1211-1218. [doi: 10.1093/ndt/gfy361] [Medline: 30541108]
38. Vart P, Grams ME, Ballew SH, Woodward M, Coresh J, Matsushita K. Socioeconomic status and risk of kidney dysfunction: the Atherosclerosis Risk in Communities study. Nephrol Dial Transplant 2019 Aug 01;34(8):1361-1368. [doi: 10.1093/ndt/gfy142] [Medline: 29897587]

Abbreviations

AI: artificial intelligence
ANCOVA: analysis of covariance
CKD: chronic kidney disease
eGFR: estimated glomerular filtration rate
ESRD: end-stage renal disease
KDQOL-SF: Short Form Kidney Disease Quality of Life
mHealth: mobile health

Edited by G Eysenbach; submitted 20.04.20; peer-reviewed by MA Bahrami, A Sudaryanto, T Svensson; comments to author 11.05.20; revised version received 12.09.20; accepted 11.11.20; published 15.12.20.

Please cite as:
Li WY, Chiu FC, Zeng JK, Li YW, Huang SH, Yeh HC, Cheng BW, Yang FJ
Mobile Health App With Social Media to Support Self-Management for Patients With Chronic Kidney Disease: Prospective Randomized Controlled Study
J Med Internet Res 2020;22(12):e19452
URL: http://www.jmir.org/2020/12/e19452/
doi:10.2196/19452
PMID:33320101

Original Paper

Prostate Cancer Risk Calculator Apps in a Taiwanese Population Cohort: Validation Study

I-Hsuan Alan Chen[1,2,3,4], MSc, MD; Chi-Hsiang Chu[5], PhD; Jen-Tai Lin[1,2], MD, PhD; Jeng-Yu Tsai[1,2], MD, PhD; Chia-Cheng Yu[1,2,3], MD; Ashwin Narasimha Sridhar[4,6], FRCS, MSc; Prasanna Sooriakumaran[4,6,7], FRCS, PhD; Rui C V Loureiro[4,8], PhD; Manish Chand[4,9], MBA, FRCS, PhD

[1]Division of Urology, Department of Surgery, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan

[2]School of Medicine, National Yang-Ming University, Taipei, Taiwan

[3]Division of Urology, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

[4]Division of Surgery and Interventional Sciences, University College London, London, United Kingdom

[5]Department of Statistics, National Cheng Kung University, Tainan, Taiwan

[6]Department of Uro-Oncology, University College London Hospital, London, United Kingdom

[7]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom

[8]Aspire Centre for Rehabilitation Engineering and Assistive Technology, University College London and the Royal National Orthopaedic Hospital, Stanmore, United Kingdom

[9]Department of Colorectal Surgery, University College London Hospital, London, United Kingdom

Corresponding Author:
Jen-Tai Lin, MD, PhD
Division of Urology
Department of Surgery
Kaohsiung Veterans General Hospital
No 386, Dazhong 1st Rd
Zuoying District
Kaohsiung, 813414
Taiwan
Phone: 886 975939290
Email: jtlin@vghks.gov.tw

Abstract

Background: Mobile health apps have emerged as useful tools for patients and clinicians alike, sharing health information or assisting in clinical decision-making. Prostate cancer (PCa) risk calculator mobile apps have been introduced to assess risks of PCa and high-grade PCa (Gleason score ≥7). The Rotterdam Prostate Cancer Risk Calculator and Coral–Prostate Cancer Nomogram Calculator apps were developed from the 2 most-studied PCa risk calculators, the European Randomized Study of Screening for Prostate Cancer (ERSPC) and the North American Prostate Cancer Prevention Trial (PCPT) risk calculators, respectively. A systematic review has indicated that the Rotterdam and Coral apps perform best during the prebiopsy stage. However, the epidemiology of PCa varies among different populations, and therefore, the applicability of these apps in a Taiwanese population needs to be evaluated. This study is the first to validate the PCa risk calculator apps with both biopsy and prostatectomy cohorts in Taiwan.

Objective: The study's objective is to validate the PCa risk calculator apps using a Taiwanese cohort of patients. Additionally, we aim to utilize postprostatectomy pathology outcomes to assess the accuracy of both apps with regard to high-grade PCa.

Methods: All male patients who had undergone transrectal ultrasound prostate biopsies in a single Taiwanese tertiary medical center from 2012 to 2018 were identified retrospectively. The probabilities of PCa and high-grade PCa were calculated utilizing the Rotterdam and Coral apps, and compared with biopsy and prostatectomy results. Calibration was graphically evaluated with the Hosmer-Lemeshow goodness-of-fit test. Discrimination was analyzed utilizing the area under the receiver operating characteristic curve (AUC). Decision curve analysis was performed for clinical utility.

Results: Of 1134 patients, 246 (21.7%) were diagnosed with PCa; of these 246 patients, 155 (63%) had high-grade PCa, according to the biopsy results. After confirmation with prostatectomy pathological outcomes, 47.2% (25/53) of patients were upgraded to high-grade PCa, and 1.2% (1/84) of patients were downgraded to low-grade PCa. Only the Rotterdam app demonstrated good calibration for detecting high-grade PCa in the biopsy cohort. The discriminative ability for both PCa (AUC: 0.779 vs 0.687; DeLong's method: *P*<.001) and high-grade PCa (AUC: 0.862 vs 0.758; *P*<.001) was significantly better for the Rotterdam app. In the prostatectomy cohort, there was no significant difference between both apps (AUC: 0.857 vs 0.777; *P*=.128).

Conclusions: The Rotterdam and Coral apps can be applied to the Taiwanese cohort with accuracy. The Rotterdam app outperformed the Coral app in the prediction of PCa and high-grade PCa. Despite the small size of the prostatectomy cohort, both apps, to some extent, demonstrated the predictive capacity for true high-grade PCa, confirmed by the whole prostate specimen. Following our external validation, the Rotterdam app might be a good alternative to help detect PCa and high-grade PCa for Taiwanese men.

(J Med Internet Res 2020;22(12):e16322) doi:10.2196/16322

KEYWORDS

diagnosis; mHealth; mobile apps; prostate cancer; prostate-specific antigen; risk calculator

Introduction

The use of health-related apps is increasing within health care systems. Prostate cancer (PCa) risk calculator mobile apps have been introduced to assess risks of PCa and high-grade PCa (Gleason score ≥7). The Rotterdam Prostate Cancer Risk Calculator and Coral–Prostate Cancer Nomogram Calculator apps were developed from the 2 most-studied PCa risk calculators, the European Randomized Study of Screening for Prostate Cancer (ERSPC) [1] and the North American Prostate Cancer Prevention Trial (PCPT) [2] risk calculators, respectively. Adam et al [3] performed a critical appraisal of 7 PCa risk calculator apps, indicating that the Rotterdam and Coral apps performed best during the prebiopsy stage. According to the currently available evidence, both apps have only been externally validated by a 2-center European study. They have demonstrated better predictive accuracy than prostate-specific antigen (PSA) and digital rectal examination (DRE) [4].

In Taiwan, around 40% of new PCa cases are diagnosed as locally advanced or metastatic diseases, which is less favorable than the stage distribution of Western countries [5]. This has not changed remarkably over the last two decades, albeit the incidence of PCa has been increasing in Taiwan since 1979. Metastatic PCa still made up almost 30% of newly-diagnosed cases from 2004 to 2012 compared to a proportion of 32.7% from 1977 to 1997 [6]. For early detection of PCa, risk calculator apps may help assess the risk of PCa or high-grade PCa in the Taiwanese population. Moreover, with the capacity to differentiate high-grade PCa, active surveillance might be supported by these apps during patient counseling.

The aim of this study is to evaluate the performance of PCa risk calculator apps in a Taiwanese population. We performed external validation using a Taiwanese cohort of patients who had undergone transrectal ultrasound (TRUS) prostate biopsy. Additionally, in previous validation studies for PCa risk calculators or apps, risk stratification was based on biopsy outcomes instead of postprostatectomy pathology results. Accordingly, we aimed to utilize postprostatectomy pathology outcomes to assess the accuracy of both apps with regard to high-grade PCa.

Methods

Inclusion Criteria

Internal review board approval (IRB No.: VGHKS19-CT3-13) was granted by a Taiwanese tertiary medical center, the Kaohsiung Veterans General Hospital. All male patients (N=1344) undergoing TRUS prostate biopsies with a 12-core systematic biopsy strategy from 2012 to 2018 were enrolled. The indication for prostate biopsy included an abnormal PSA level (>4 ng/mL) or an abnormal DRE. Each patient would receive DRE and TRUS before the biopsy was performed; prostate volume (PV) was calculated by the ellipsoid formula (length x width x height x π/6). Some patients (53/1344) had multiparametric magnetic resonance imaging (mpMRI) scans on a self-pay basis because the Taiwan National Health Insurance system has not approved the reimbursement of pelvic magnetic resonance imaging (MRI) before prostate biopsy. All prebiopsy mpMRI scans were reported by dedicated urologic radiologists, in agreement with the Prostate Imaging Reporting and Data System, version 2 (PI-RADS v2) [7]. The number of patients who received radical prostatectomy was 137. Consultant pathologists reviewed all biopsies and postprostatectomy specimens.

According to the Rotterdam app, the definition of clinically significant PCa is a tumor stage greater than T2b, or a Gleason biopsy score of ≥7, which is identical to high-risk PCa in the ERSPC risk calculator (ERSPC-RC) [8]. In comparison, the Coral app defined a Gleason biopsy score of ≥7 as high-grade PCa, which originated from the PCPT [2]. In order to use consistent terminology, a Gleason score of ≥7 was defined as high-grade PCa. PSA was designated as the latest total serum PSA level before prostate biopsy.

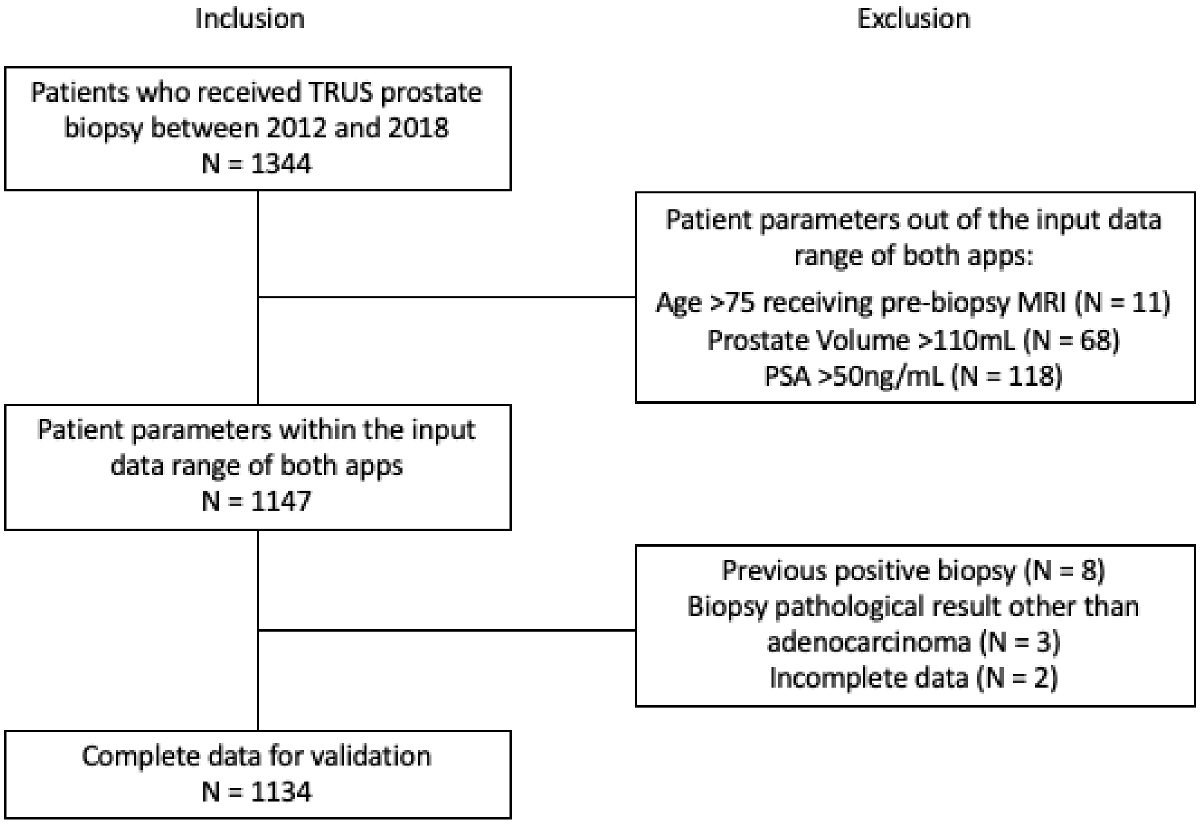
Data Collection

All patient data were retrospectively collected via electronic medical records. The Rotterdam app accepts input data on age, DRE history and outcome, previous negative biopsy, PSA, PV, volume measure method (TRUS or DRE), TRUS evaluation (normal or abnormal), MRI history, and PI-RADS score. The Coral app requires data on ethnicity (African American, Caucasian, Hispanic, or Other), age, DRE, PSA, family history, and prior biopsy results. Following input data collection, the risks of PCa and high-grade PCa were calculated using the Rotterdam and Coral apps.

Exclusion Criteria

Each app customizes its parameters to impose controls or constraints on accepted input values; that is, both the Rotterdam and Coral apps have an input data range for some parameters. For instance, the input PSA range is limited from 0.4-50 ng/mL within the Rotterdam app and 0.3-100 ng/mL within the Coral app. If the patients have prebiopsy MRI scans, there is an age limit between 50 and 75 years. On the contrary, without prebiopsy MRI, both the Rotterdam and Coral apps have no limitation on the input age data. According to the accepted input values from both apps, 197 patients were excluded, either because their PV was <10 or >110 mL (68/197), or their PSA level was <0.4 or >50 ng/mL (118/197), or they underwent prebiopsy MRI at the age of <50 or >75 years (11/197). In addition, 13 more patients with previous positive biopsies (8/13), pathological diagnosis different from adenocarcinoma (3/13), or incomplete data (2/13) were excluded. Details of the inclusion and exclusion process are illustrated in the flow chart in Figure 1.

Figure 1. Flowchart of patient enrollment into the study. MRI: magnetic resonance imaging; PSA: prostate-specific antigen; TRUS: transrectal ultrasound.



Statistical Analysis

Statistical analyses were performed utilizing SPSS (version 18; IBM Corp) and R software packages (R Core Team). The Kolmogorov-Smirnov test was used to examine the normality of the distribution of variables. Categorical variables were assessed with the chi-square test. Continuous variables were described as medians and interquartile ranges or means and standard deviations, and compared by the Mann-Whitney *U* test or the Student independent *t* test based upon their nonnormal or normal distributions, respectively. The applicability of each PCa risk calculator app in the Taiwanese population cohort was statistically analyzed on the basis of its discrimination, calibration, and clinical utility [9]. Calibration relates to the agreement between the observed and predicted proportion of events; calibration was evaluated graphically utilizing a calibration plot in which the observed probabilities were plotted against the predicted probabilities, enabling assessment of the extent of risk underestimation or overestimation [10]. The statistical significance of miscalibration was examined by the Hosmer-Lemeshow goodness-of-fit test [11].

Discrimination reflects the capacity of a prediction model to differentiate between those with and without an event (any-grade or high-grade PCa) and is quantified utilizing the area under the receiver operating characteristic (ROC) curve (AUC). The AUCs of the Rotterdam and Coral apps were compared using DeLong’s method [12]. As for clinical utility, decision curve analysis was performed to analyze whether both apps were beneficial for clinical decision-making or which app would lead to better decisions. We calculated the net benefit to quantify the clinical utility; different threshold probabilities mean different harm-to-benefit ratios. Net benefit was formulated as the number of true positives subtracted from the proportion of false positives weighted by the odds of the risk threshold probability, and the result was divided by the sample size. By measuring the proportions of *net* true positives in the models, we could assess whether any model performed better than others and the default strategies of biopsying all or no patients across the reasonable range of risk threshold probabilities [13].

Results

Patient Demographics

Of 1344 patients undergoing biopsies, 246 (21.7%) patients were diagnosed with PCa; of these 246 patients, 155 (63%) had high-grade PCa, according to the biopsy results. Compared to males with negative biopsies, patients with PCa were significantly older, had higher PSA levels, smaller PVs, more abnormal findings on TRUS and DRE, and higher PI-RADS scores demonstrated on mpMRI (Table 1). Both in the biopsy and prostatectomy cohorts, patients with high-grade PCa had significantly higher PSA, lower PV, and more abnormal findings on TRUS compared to those with low-grade PCa (Table 2). Among 246 diagnoses of PCa, 137 patients underwent radical prostatectomy; based on the postprostatectomy outcomes, 47.2% (25/53) of patients were upgraded to high-grade PCa and 1.2% (1/84) of patients were downgraded to low-grade PCa.

Table 1. Patient demographics (N=1344). Categorical variables were assessed with the chi-square test; continuous variables were compared by the Mann-Whitney U test or the Student independent t test based on their nonnormal or normal distribution, respectively.

Characteristics	All patients (n=1134)	Patients with no cancer (n=888, 78.3%)	Patients with cancer (n=246, 21.7%)	*P* value
Age in years, mean (SD); median (1st quartile-3rd quartile)	68.78 (8.23); 67 (61-73)	66.31 (8.17); 66 (61-72)	68.49 (8.25); 69 (63-74)	<.001
PSA[a], mean (SD); median (1st quartile-3rd quartile)	10.27 (7.44); 8.08 (5.73-11.65)	9.20 (5.90); 7.66 (5.57-10.44)	14.12 (10.53); 10.16 (6.64-17.51)	<.001
PV[b], mean (SD); median (1st quartile-3rd quartile)	52.72 (21.13); 48.35 (37.08-66.01)	55.35 (20.86); 51.52 (39.93-68.88)	43.25 (19.36), 37.82 (29.68-52.00)	<.001
Family history, n (%)	44 (3.9)	32 (3.6)	12 (4.9)	.360
Suspicious TRUS[c], n (%)	253 (22.3)	151 (17.0)	102 (41.5)	<.001
Suspicious DRE[d], n (%)	192 (16.9)	89 (10.0)	103 (41.9)	<.001
MRI[e] (n=38) PI-RADS[f] 4,5, n (%)	27 (71.1)	14 (58.3)	13 (92.9)	.030
Rotterdam PCa[g](%), mean (SD); median (1st quartile-3rd quartile)	29.98 (23.98); 21 (13-39)	23.92 (17.69); 18 (12-31)	51.84 (30.20); 50 (24-84)	<.001
Rotterdam high-grade PCa (%), mean (SD); median (1st quartile-3rd quartile)	14.30 (21.47); 5 (2-14)	8.66 (12.64); 4 (2-9)	34.60 (31.98); 21 (6-65)	<.001
Coral PCa (%), mean (SD); median (1st quartile-3rd quartile)	34.37 (11.53); 32 (26-39)	32.42 (9.64); 31 (26-36)	41.42 (14.63); 38 (31.00-51.25)	<.001
Coral high-grade PCa (%), mean (SD); median (1st quartile-3rd quartile)	14.59 (10.38); 11 (8-18)	12.74 (8.05); 11 (7-16)	21.24 (14.38); 17 (10-30)	<.001

[a] PSA: prostate-specific antigen.

[b] PV: prostate volume.

[c] TRUS: transrectal ultrasound.

[d] DRE: digital rectal examination.

[e] MRI: magnetic resonance imaging.

[f] PI-RADS: Prostate Imaging Reporting and Data System.

[g] PCa: prostate cancer.

Table 2. Demographics of patients with prostate cancer (PCa; n=383). Categorical variables were assessed with the chi-square test; continuous variables were compared by the Mann-Whitney U test or the Student independent t test based on their nonnormal or normal distribution, respectively.

Characteristics	Biopsy cohort (n=246)			Prostatectomy cohort (n=137)		
	Low-grade PCa[a] (n=91, 37%)	High-grade PCa (n=155, 63%)	*P* value	Low-grade PCa (n=29, 21%)	High-grade PCa (n=108, 79%)	*P* value
Age in years, mean (SD)	67.22 (8.00)	69.24 (8.33)	.064	63.07 (7.21)	65.79 (6.27)	.047
PSA[b], mean (SD); median (1st quartile-3rd quartile)	9.94 (7.36); 7.9 (5.1-12.2)	16.57 (11.33); 13.0 (8.1-23.2)	<.001	6.68 (2.60); 7.1 (4.7-8.4)	14.33 (10.27); 11.1 (6.7-17.9)	<.001
PV[c], mean (SD); median (1st quartile-3rd quartile)	47.56 (22.30); 42.4 (30.4-59.7)	40.71(16.97); 36.4 (28.9-48.9)	.039	51.53 (22.13); 46.6 (32.5-61.0)	39.80 (17.26); 34.2 (28.6-46.3)	.004
Family history, n (%)	4 (4.4)	8 (5.2)	>.99	1 (3.4)	7 (6.5)	>.99
Suspicious TRUS[d], n (%)	18 (19.8)	84 (54.2)	<.001	1 (3.4)	41 (38.0)	<.001
Suspicious DRE[e], n (%)	27 (29.7)	76 (49.0)	.003	7 (24.1)	45 (41.7)	.084
MRI[f] (n=14) PI-RADS[g] 4,5, n (%)	2 (100.0)	11 (91.7)	>.99	-	10 (90.9)	-
Rotterdam PCa (%), mean (SD); median (1st quartile-3rd quartile)	35.13 (26.39); 25.0 (15.0-54.0)	61.65 (27.96); 66.0 (36.0-88.0)	<.001	21.72 (14.70); 18.0 (12.5-25.0)	54.32 (27.91); 52.0 (28.3-84.0)	<.001
Rotterdam high-grade PCa (%), mean (SD); median (1st quartile-3rd quartile)	18.74 (25.72); 7.0 (3.0-21.0)	43.92 (31.69); 36.0 (15.0-75.0)	<.001	7.14 (8.50); 4.0 (3.0-7.0)	34.39 (28.93); 23.5 (10.0-60.3)	<.001
Coral PCa (%), mean (SD); median (1st quartile-3rd quartile)	35.11 (11.86); 31.0 (26.0-41.0)	45.13 (14.86); 42.0 (34.0-55.0)	<.001	28.34 (6.14); 27.0 (24.0-34.0)	39.43 (12.29); 37.0 (30.0-48.0)	<.001
Coral high-grade PCa (%), mean (SD); median (1st quartile-3rd quartile)	15.37 (10.71); 12.0 (8.0-19.0)	24.68 (15.16); 20.0 (13.0-34.0)	<.001	9.48 (4.99); 8.0 (6.0-13.5)	18.92 (11.55); 15.0 (10.0-25.0)	<.001

[a] PCa: prostate cancer.

[b] PSA: prostate-specific antigen.

[c] PV: prostate volume.

[d] TRUS: transrectal ultrasound.

[e] DRE: digital rectal examination.

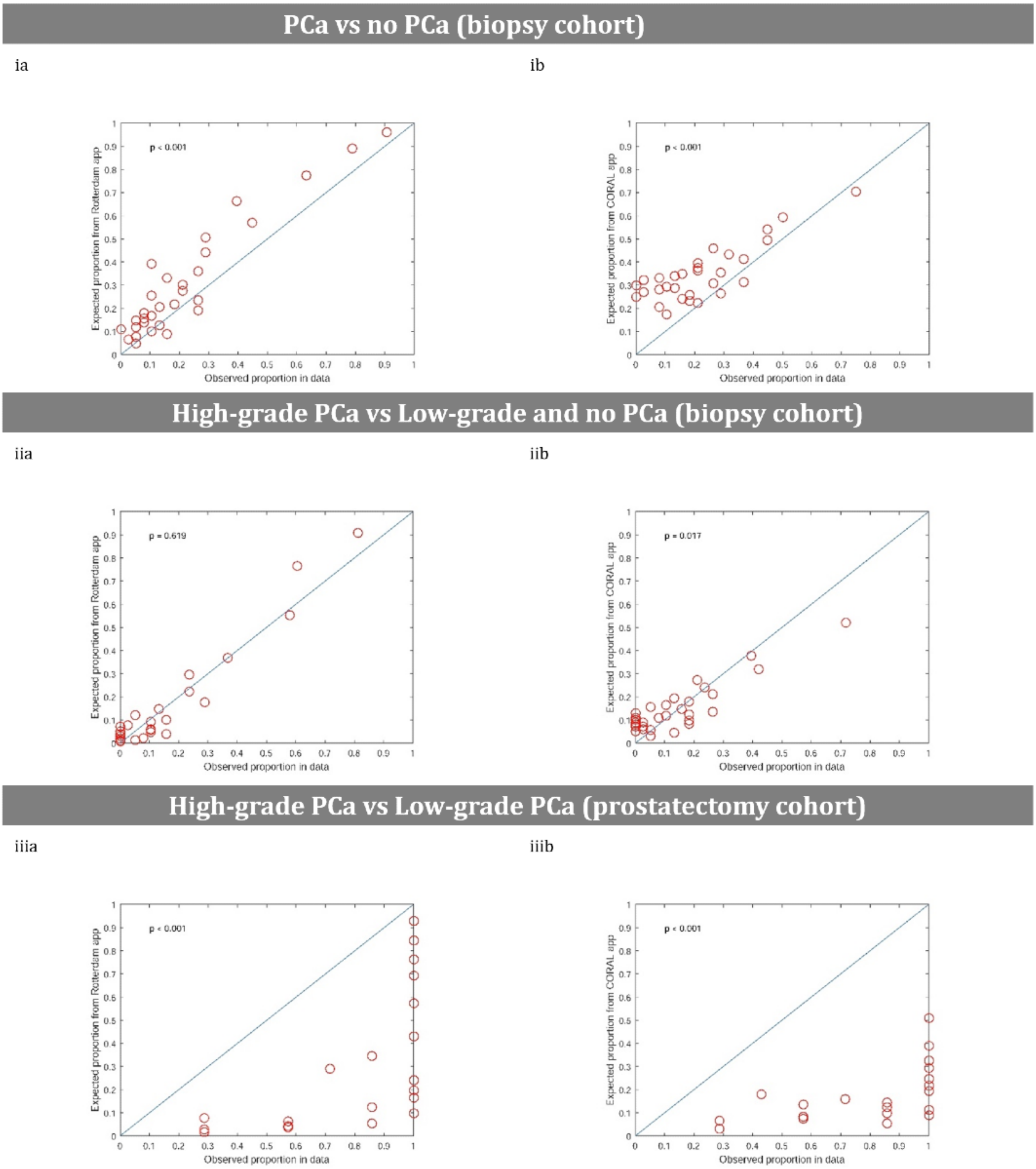
[f] MRI: magnetic resonance imaging.

[g] PI-RADS: Prostate Imaging Reporting and Data System.

Calibration

The calibration of both apps was tested with the Hosmer-Lemeshow goodness-of-fit test (Figure 2). Comparing both apps, only the Rotterdam app demonstrated a good calibration (*P*=.619) for detecting high-grade PCa in the biopsy cohort. Other models were miscalibrated, including all models created from the Coral app.

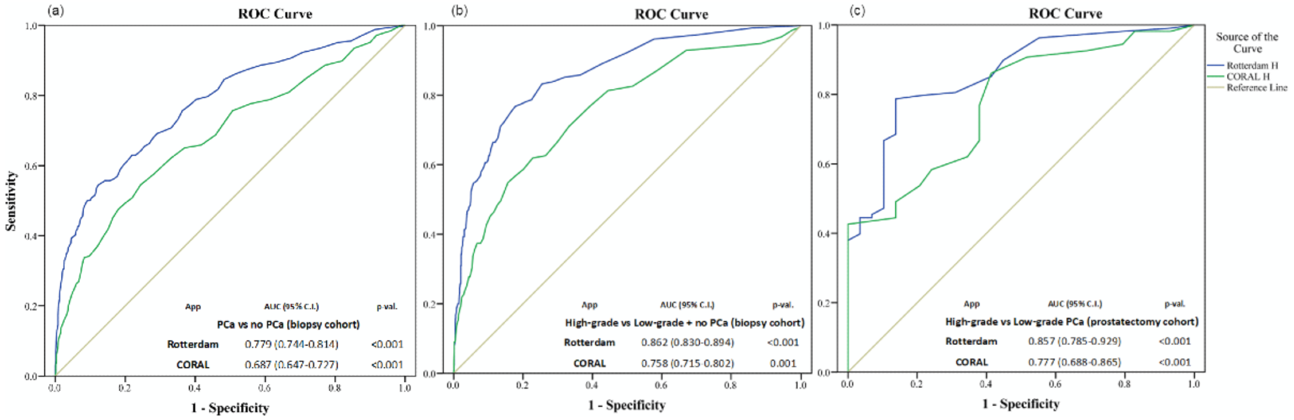
Figure 2. Calibration plots comparing (a) the Rotterdam app and (b) Coral app showing the agreement between (i) predicted and observed probabilities for diagnosing prostate cancer (PCa) and (ii) high-grade PCa in the biopsy cohort, and (iii) high-grade PCa in the prostatectomy cohort. Each circle in the plots represents a group of patients with an observed probability of PCa or high-grade PCa on the x-axis, corresponding to an average calculated risk of PCa or high-grade PCa by the apps on the y-axis. Figures (ia) and (ib) demonstrated overestimation, whereas Figures (iiia) and (iiib) illustrated underestimation. Only Figure (iia) showed a good calibration; in Figure (iib), overestimation was revealed among the lower observed proportions and underestimation among the higher observed proportions.



Discrimination

Both the Rotterdam and Coral apps could significantly predict PCa and high-grade PCa in the biopsy cohort on ROC analysis (Figure 3, a and b). The discriminative capacity for detection of both PCa (AUC: 0.779 vs 0.687; DeLong's method: *P*<.001) and high-grade PCa (AUC: 0.862 vs 0.758; *P*<.001) was significantly better for the Rotterdam app compared to the Coral app. In the prostatectomy cohort, the Rotterdam and Coral apps were not significantly different for predicting high-grade PCa (AUC: 0.857 vs 0.777; *P*=.128; Figure 3c).

Figure 3. Receiver operating characteristic (ROC) curves and areas under the receiver operating characteristic curves (AUCs) for the discriminative ability of the Rotterdam and Coral apps. (a) Prostate cancer (PCa) vs no PCa in the biopsy cohort; (b) high-grade PCa vs low-grade PCa plus no PCa in the biopsy cohort; (c) high-grade PCa vs low-grade PCa in the prostatectomy cohort.



Clinical Utility

In the decision curve analysis for the biopsy cohort of patients, both apps demonstrated clinical net benefits in the threshold probability range of 10% to 85% for the detection of any PCa. In the detection of high-grade PCa, the Rotterdam and Coral apps provided net benefits in the threshold probability range of 5%-70% and 10%-80%, respectively. In comparing both apps, the net benefit was greater for the Rotterdam app in the prediction of both PCa and high-grade PCa across the range of threshold probabilities from 5%-70% (Figure 4). It seemed that both apps provided net benefits for the detection of high-grade PCa in the prostatectomy cohort (Figure 5). Nevertheless, the prevalence of high-grade PCa in the prostatectomy cohort was very high (79%). With baseline risk being very high, it would be difficult for both apps to push the risk to a level low enough for advice against biopsy. Both apps had higher net benefits when the curves diverged at the threshold probability of about 50%, and therefore, both apps lacked value for the prediction of high-grade PCa in the prostatectomy cohort.

Figure 4. Decision curve analyses quantifying clinical utility by showing the net benefits associated with the use of the Rotterdam app (blue line) and the Coral app (red line) in (a) the detection of prostate cancer (PCa) and (b) high-grade PCa. Decision curves investigate the theoretical net benefit at various threshold probabilities. The oblique gray line assumes that all persons will undergo prostate biopsy, whereas the horizontal black line along the x-axis assumes that no one will receive biopsy. The threshold probability may correspond to the calculated prostate cancer risk. The area under the curve between these 2 lines illustrates net benefits. In the same range of threshold probability, higher net benefits represent better clinical utility.

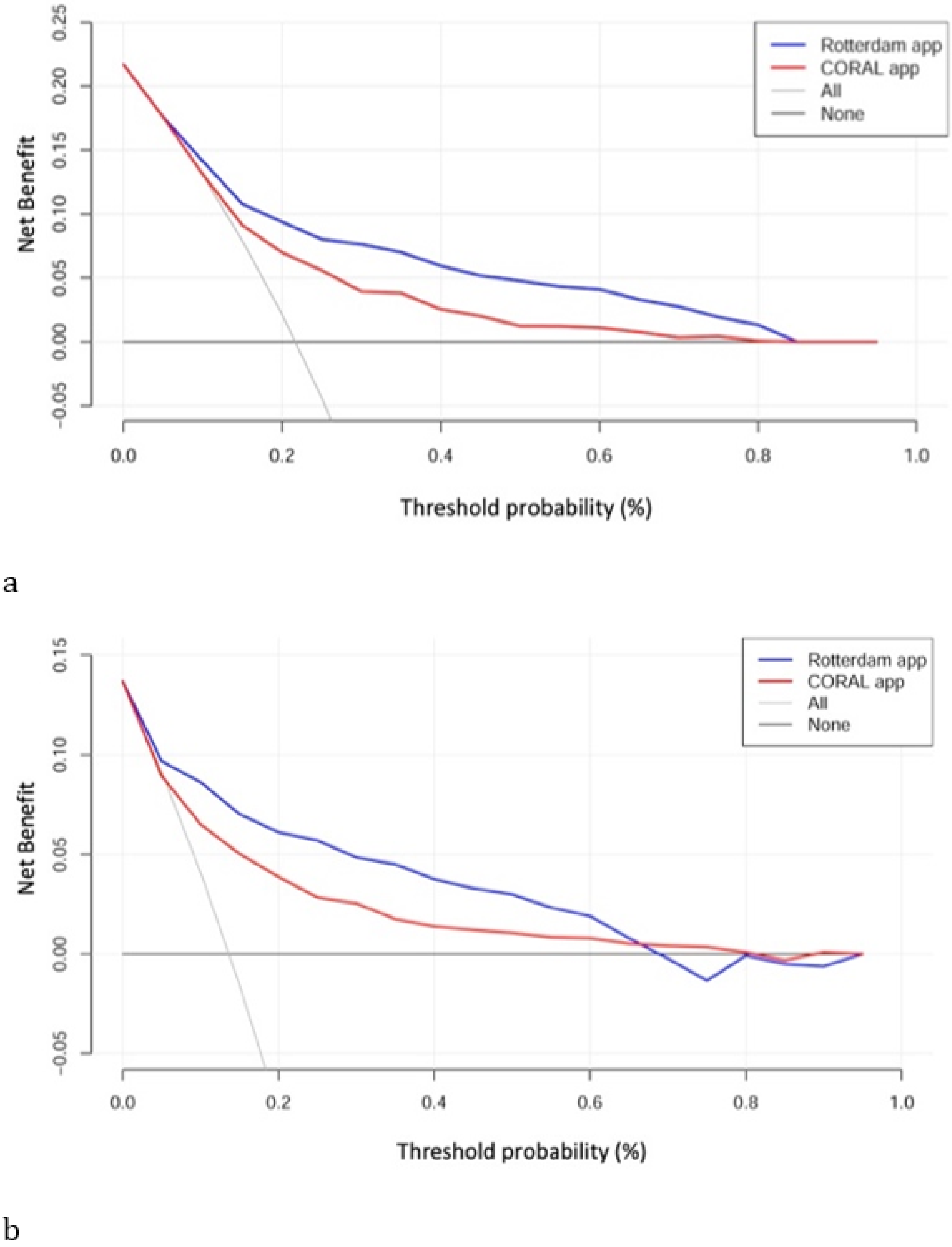
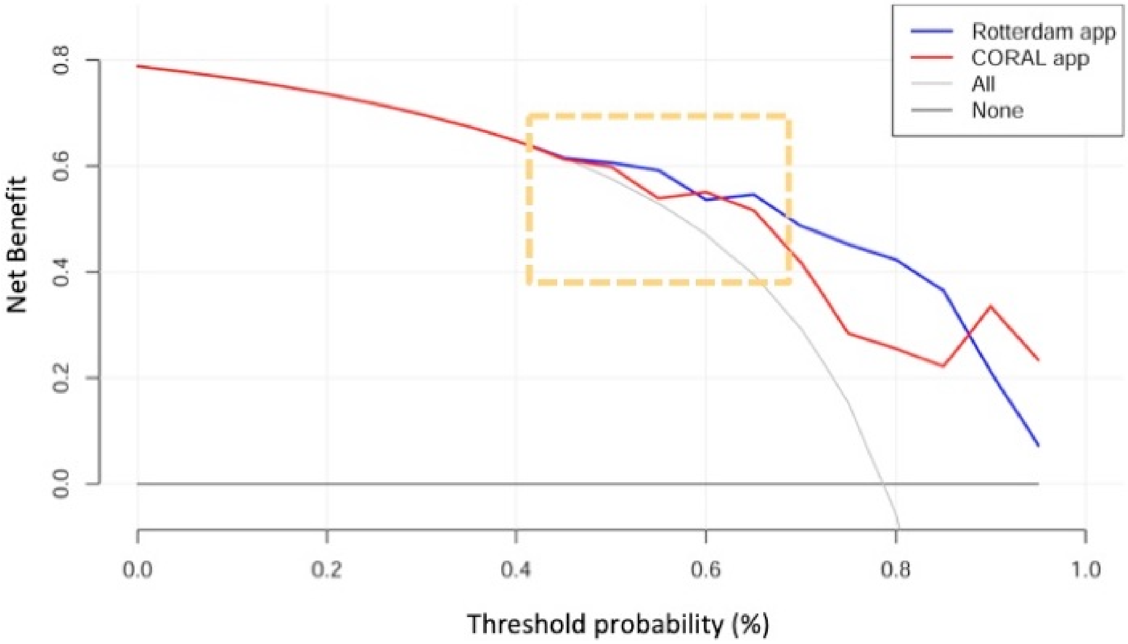


Figure 5. Decision curve analysis showing the net benefits of utilizing the Rotterdam app (blue line) and the Coral app (red line) to detect high-grade prostate cancer (PCa) in the prostatectomy cohort. The curves are skewed because the incidence of high-grade PCa in the prostatectomy cohort is relatively high and the sample size is small. No net benefit could be demonstrated below the risk threshold of 50%.



Discussion

Principal Findings

In this study, we found that the Rotterdam and Coral apps were both applicable to the Taiwanese cohort of patients who had undergone TRUS prostate biopsy, even though these apps were built based on Western populations. In order to externally validate these 2 apps, 3 key statistical measures were used in the assessment of predictive performance. Firstly, most models revealed miscalibration, but the Rotterdam app demonstrated good calibration for the prediction of high-grade PCa in the biopsy cohort. Secondly, the Rotterdam app outperformed the Coral app in its discriminative ability for predicting PCa and high-grade PCa in the biopsy cohort. Thirdly, the Rotterdam app provided greater net benefits than the Coral app to assist in biopsy decision-making. In brief, the Rotterdam app delivered better predictive performance than the Coral app for PCa and high-grade PCa in our Taiwanese population cohort.

To the best of our knowledge, at the stage before PCa is diagnosed, no risk prediction model has ever been validated by the whole prostate specimen. Data from the 137 patients who had undergone radical prostatectomy for any-grade PCa were utilized to evaluate the Rotterdam and Coral apps' predictive capacity for high-grade PCa. Both apps demonstrated fairly good discrimination for predicting high-grade PCa in the prostatectomy cohort and the biopsy cohort. It was implied that after confirmation with postprostatectomy pathology outcomes, both apps still delivered a comparable discriminative ability for predicting high-grade PCa in the present Taiwanese cohort. However, during calibration and decision curve analysis in the prostatectomy cohort, both apps were miscalibrated and revealed few net benefits. This might be explained by the small sample size and different pathology distribution, as there was a higher prevalence of high-grade PCa (79%). Moreover, these apps were built on biopsy cohorts, which are different from the prostatectomy cohort. More patients with radical prostatectomy might need to be enrolled to validate app predictability of high-grade PCa, which would be validated by the whole prostate specimen.

In current clinical practice, most patients with an abnormal PSA level of >4 ng/mL or an abnormal DRE are put forward for biopsy. However, such indications lead to a myriad of unnecessary biopsies and associated complications, such as hematuria, hematospermia, rectal bleeding, acute urinary retention, urinary tract infection, or even sepsis. To increase the accuracy of cancer detection and reduce unnecessary postbiopsy morbidities, several biomarker tests have been developed, including the Prostate Health Index (PHI), percent-free PSA, PCA3, 4K-score, etc [14]. The European Association of Urology (EAU) guidelines suggest an individualized evaluation of PCa risk. Age, family history of PCa, DRE, serum or urine markers, and mpMRI are validated parameters in combination with PSA levels to help predict the risk of PCa [15]. Recently, professionals have formulated a number of PCa risk calculators using some of these useful predictors to improve predictive accuracy, and such multivariable risk approaches have performed better than PSA or DRE alone [10]. Nonetheless, most of them have only been validated in independent cohorts; neither superiority nor global applicability has been shown [16]. De Nunzio et al [4] had validated the Rotterdam and Coral apps' discriminative abilities utilizing a southern European cohort as providing better predictive performance than PSA or DRE; however, the predictability of PCa or high-grade PCa in the Taiwanese population stills needs to be addressed.

It is well known that Gleason upgrading occurs in 32%-49% of patients with initial biopsy of low-grade (Gleason 3+3) PCa at the time of pathological assessment of the whole prostate specimen [14]. Verep et al [17] reviewed 137 patients who were eligible for active surveillance but underwent radical prostatectomy at their institution. The criteria of active surveillance included Gleason 3+3 adenocarcinoma, maximum 2 positive biopsy cores, PSA <10ng/mL, and clinical T-stage equal or less than 2a. Following pathological confirmation, Gleason upgrading was noted in almost half of the patients (49.3%), and upstaging to pT3a occurred in 17 patients (12.5%) [17]. Due to the risks of over-diagnosis and over-treatment for clinically insignificant PCa, active surveillance has become increasingly adopted as a preferred treatment option for patients with low-grade PCa. However, without precise risk stratification, active surveillance might delay the timing of curative treatment for localized PCa, or even increase the risks of lymph node involvement and distant metastasis. Consequently, the accuracy of risk prediction tools has become of paramount importance.

Mobile health (mHealth) is regarded as a valuable tool to implement patient-centered care, which is in accordance with the individualized risk assessment of PCa recommended by the EAU guidelines. mHealth can provide access to health information, skills, and services, and can also promote positive health behavioral changes to prevent acute and chronic diseases. Real-time monitoring can obtain live data from patients and transmit inputs to a network or a medical app on a smartphone to assist clinical decision-making. Regardless of the environmental circumstances, geographical barriers, and conventional infrastructures, it can share timely information between patients and health personnel, replacing the traditional face-to-face platform of medical care. Nevertheless, to not harm patients, it is pivotal that scientific accuracy, patient safety, and user privacy of mHealth apps be assured [18].

One systematic review that critically appraised PCa risk calculator apps maintained that the Rotterdam Prostate Cancer Risk Calculator and Coral–Prostate Cancer Nomogram Calculator outperformed other apps [3]. The authors utilized the validated user version of the Mobile Application Rating Scale to individually assess and rate 7 apps, including 3 categories of app quality ratings, subjective quality, and perceived impact [19]. Objective characteristics were thoroughly documented and assessed. None of these apps allowed confidentiality, data storage developing trends, or customization. Both the Rotterdam and Coral apps were found to help differentiate low-grade from high-grade PCa, a noteworthy characteristic of patient counseling in active surveillance compared to other curative alternatives.

Mobile technology enables clinicians and patients to download the Rotterdam and Coral apps readily, and the owners can utilize these apps without an internet connection. They both have the advantages of being less time-consuming and more cost-effective, delivering better immediacy, upgradability, and shareability than the original risk calculators. In addition, they are globally available and recommended by the American Urological Association and EAU guidelines to improve prediction and help determine the risk of PCa stratification. Compared with the Coral app, the conspicuous disadvantage of the Rotterdam app is its cost ($1.99 USD). However, it has been proven to reach a wider audience, with availability on both Apple and Android platforms. While the Coral app has merely 1 language choice, the Rotterdam app has 7 different language options, including Chinese, Dutch, English, Estonian, German, Portuguese, and Spanish.

Limitations

This study was a single-institution retrospective study, and more cohorts from other Taiwanese hospitals are required to confirm the results. Also, the sample size in the prostatectomy cohort was rather small (n= 137). Notwithstanding that, this is the first study validating PCa risk calculators with postprostatectomy pathological outcomes, although the predictive accuracy of both apps for true high-grade PCa could not be completely determined. Further, no PHI data are available in our institution. The PHI is a combination of 3 blood tests measuring different forms of the PSA protein (total PSA, free PSA, and p2PSA) and calculated as (p2PSA/fPSA) × √tPSA. It is one of the predictors listed in the Rotterdam app, although the risks of PCa can still be calculated without PHI data. Few patients underwent mpMRI before biopsy in our cohort; however, mpMRI has emerged as an important prediction tool to identify clinically significant PCa, especially before a repeated biopsy, which has been recommended by the guidelines [14,15]. The Rotterdam app and web-based ERSPC-RC is one of the PCa risk prediction models incorporating mpMRI (PI-RADS 1-5) [20]; nevertheless, the predictor of the PI-RADS used in the Rotterdam app was the first version, and most of our MRI images were graded according to the PI-RADS v2 guidelines. Moreover, the issues of interobserver variability and heterogeneous definitions of *abnormality* in mpMRI interpretation remain to be explored, and these MRI risk prediction models need to be validated further.

Conclusions

In our external validation study, the Rotterdam and Coral mHealth apps could be applied to the Taiwanese cohort of patients. Following an assessment of calibration, discrimination, and clinical utility, the Rotterdam app outperformed the Coral app for predicting both any-grade PCa and high-grade PCa. The size of the prostatectomy cohort was small; however, both mobile phone apps, to some extent, demonstrated a predictive capacity for true high-grade PCa, confirmed by the whole prostate specimen. As of yet, there is no PCa risk calculator app developed specifically for the Taiwanese population; however, the Rotterdam app might be a good alternative to enhance the predictive accuracy of current methods for detecting PCa and high-grade PCa.

Conflicts of Interest

None declared.

References

1. Roobol MJ, Verbeek JF, van der Kwast T, Kümmerlin IP, Kweldam CF, van Leenders GJ. Improving the Rotterdam European Randomized Study of Screening for Prostate Cancer Risk Calculator for Initial Prostate Biopsy by Incorporating the 2014 International Society of Urological Pathology Gleason Grading and Cribriform growth. European Urology 2017 Jul;72(1):45-51. [doi: 10.1016/j.eururo.2017.01.033]
2. Ankerst DP, Hoefler J, Bock S, Goodman PJ, Vickers A, Hernandez J, et al. Prostate Cancer Prevention Trial Risk Calculator 2.0 for the Prediction of Low- vs High-grade Prostate Cancer. Urology 2014 Jun;83(6):1362-1368. [doi: 10.1016/j.urology.2014.02.035]
3. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. 'Prostate Cancer Risk Calculator' mobile applications (Apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2017 Dec 8;36(4):565-573. [doi: 10.1007/s00345-017-2150-1]
4. De Nunzio C, Lombardo R, Tema G, Cancrini F, Russo GI, Chacon R, et al. Mobile phone apps for the prediction of prostate cancer: External validation of the Coral and Rotterdam apps. European Journal of Surgical Oncology 2019 Mar;45(3):471-476. [doi: 10.1016/j.ejso.2018.09.018]
5. Hung C, Yang C, Ou Y. Urologic cancer in Taiwan. Jpn. J. Clin. Oncol 2016 Apr 06;46(7):605-609. [doi: 10.1093/jjco/hyw038]
6. Pu Y. Prostate cancer in Taiwan: epidemiology and risk factors. Int J Androl 2000 Apr;23(S2):34-36. [doi: 10.1046/j.1365-2605.2000.00012.x]
7. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. European Urology 2016 Jan;69(1):16-40. [doi: 10.1016/j.eururo.2015.08.052]
8. Pereira-Azevedo N, Osório L, Fraga A, Roobol MJ. Rotterdam Prostate Cancer Risk Calculator: Development and Usability Testing of the Mobile Phone App. JMIR Cancer 2017 Jan 06;3(1):e1. [doi: 10.2196/cancer.6750]
9. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the Performance of Prediction Models. Epidemiology 2010;21(1):128-138. [doi: 10.1097/ede.0b013e3181c30fb2]
10. Poyet C, Nieboer D, Bhindi B, Kulkarni GS, Wiederkehr C, Wettstein MS, et al. Prostate cancer risk prediction using the novel versions of the European Randomised Study for Screening of Prostate Cancer (ERSPC) and Prostate Cancer Prevention Trial (PCPT) risk calculators: independent validation and comparison in a contemporary Europe. BJU Int 2015 Oct 05;117(3):401-408. [doi: 10.1111/bju.13314]
11. Hosmer DW, Lemesbow S. Goodness of fit tests for the multiple logistic regression model. Comm. in Stats - Theory & Methods 1980;9(10):1043-1069. [doi: 10.1080/03610928008827941]
12. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics 1988 Sep;44(3):837. [doi: 10.2307/2531595]
13. Van Calster B, Wynants L, Verbeek JF, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. European Urology 2018 Dec;74(6):796-804. [doi: 10.1016/j.eururo.2018.08.038]
14. National Comprehensive Cancer Network. Prostate Cancer Early Detection (Version 2.2019). 2019. URL: https://www.nccn.org/professionals/physician_gls/pdf/prostate_detection.pdf [accessed 2019-07-29]
15. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. European Urology 2017 Apr;71(4):618-629. [doi: 10.1016/j.eururo.2016.08.003]
16. Louie K, Seigneurin A, Cathcart P, Sasieni P. Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. Annals of Oncology 2015 May;26(5):848-864. [doi: 10.1093/annonc/mdu525]
17. Verep S, Erdem S, Ozluk Y, Kilicaslan I, Sanli O, Ozcan F. The pathological upgrading after radical prostatectomy in low-risk prostate cancer patients who are eligible for active surveillance: How safe is it to depend on bioptic pathology? Prostate 2019 Jul 03:1523-1529. [doi: 10.1002/pros.23873]
18. Pereira-Azevedo NM, Venderbos LDF. eHealth and mHealth in prostate cancer detection and active surveillance. Transl. Androl. Urol 2018 Feb;7(1):170-181. [doi: 10.21037/tau.2017.12.22]
19. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849]
20. Schoots IG, Roobol MJ. Multivariate risk prediction tools including MRI for individualized biopsy decision in prostate cancer diagnosis: current status and future directions. World J Urol 2019 Mar 13;38(3):517-529. [doi: 10.1007/s00345-019-02707-9]

Abbreviations

AUC: area under the receiver operating characteristic curve
EAU: European Association of Urology
ERSPC-RC: European Randomized Study of Screening for Prostate Cancer risk calculator
DRE: digital rectal examination

mHealth: mobile health
mpMRI: multiparametric magnetic resonance imaging
MRI: magnetic resonance imaging
PCa: prostate cancer
PCPT: North American Prostate Cancer Prevention Trial
PHI: Prostate Health Index
PI-RADS: Prostate Imaging Reporting and Data System
PSA: prostate-specific antigen
PV: prostate volume
ROC: receiver operating characteristic
TRUS: transrectal ultrasound

Edited by G Eysenbach; submitted 07.10.19; peer-reviewed by PH Lin, C Fernández, R Wassersug; comments to author 13.06.20; revised version received 03.07.20; accepted 11.11.20; published 18.12.20.

Please cite as:
Chen IHA, Chu CH, Lin JT, Tsai JY, Yu CC, Sridhar AN, Sooriakumaran P, Loureiro RCV, Chand M
Prostate Cancer Risk Calculator Apps in a Taiwanese Population Cohort: Validation Study
J Med Internet Res 2020;22(12):e16322
URL: http://www.jmir.org/2020/12/e16322/
doi:10.2196/16322
PMID:33337340

Original Paper

Future Mobile Device Usage, Requirements, and Expectations of Physicians in German University Hospitals: Web-Based Survey

Oliver Maassen[1,2], MSc; Sebastian Fritsch[1,2], MD, Dr; Julia Gantner[2,3], MSc; Saskia Deffge[1,2], MSc; Julian Kunze[1,2], MD, Dr; Gernot Marx[1,2], MD, Prof Dr, FRCA; Johannes Bickenbach[1,2], MD, Prof Dr

[1]Department of Intensive Care Medicine, University Hospital RWTH Aachen, Aachen, Germany

[2]SMITH Consortium of the German Medical Informatics Initiative, Leipzig, Germany

[3]Institute of Medical Statistics, Informatics and Data Science, Jena University Hospital, Jena, Germany

Corresponding Author:
Oliver Maassen, MSc
Department of Intensive Care Medicine
University Hospital RWTH Aachen
Pauwelsstrasse 30
Aachen, 52074
Germany
Phone: 49 2418080444
Email: oliver.maassen@rwth-aachen.de

Abstract

Background: The use of mobile devices in hospital care constantly increases. However, smartphones and tablets have not yet widely become official working equipment in medical care. Meanwhile, the parallel use of private and official devices in hospitals is common. Medical staff use smartphones and tablets in a growing number of ways. This mixture of devices and how they can be used is a challenge to persons in charge of defining strategies and rules for the usage of mobile devices in hospital care.

Objective: Therefore, we aimed to examine the status quo of physicians' mobile device usage and concrete requirements and their future expectations of how mobile devices can be used.

Methods: We performed a web-based survey among physicians in 8 German university hospitals from June to October 2019. The online survey was forwarded by hospital management personnel to physicians from all departments involved in patient care at the local sites.

Results: A total of 303 physicians from almost all medical fields and work experience levels completed the web-based survey. The majority regarded a tablet (211/303, 69.6%) and a smartphone (177/303, 58.4%) as the ideal devices for their operational area. In practice, physicians are still predominantly using desktop computers during their worktime (mean percentage of worktime spent on a desktop computer: 56.8%; smartphone: 12.8%; tablet: 3.6%). Today, physicians use mobile devices for basic tasks such as oral (171/303, 56.4%) and written (118/303, 38.9%) communication and to look up dosages, diagnoses, and guidelines (194/303, 64.0%). Respondents are also willing to use mobile devices for more advanced applications such as an early warning system (224/303, 73.9%) and mobile electronic health records (211/303, 69.6%). We found a significant association between the technical affinity and the preference of device in medical care ($\chi s2=53.84$, $P<.001$) showing that with increasing self-reported technical affinity, the preference for smartphones and tablets increases compared to desktop computers.

Conclusions: Physicians in German university hospitals have a high technical affinity and positive attitude toward the widespread implementation of mobile devices in clinical care. They are willing to use official mobile devices in clinical practice for basic and advanced mobile health uses. Thus, the reason for the low usage is not a lack of willingness of the potential users. Challenges that hinder the wider adoption of mobile devices might be regulatory, financial and organizational issues, and missing interoperability standards of clinical information systems, but also a shortage of areas of application in which workflows are adapted for (small) mobile devices.

(J Med Internet Res 2020;22(12):e23955) doi:10.2196/23955

KEYWORDS

mobile devices; mobile applications; apps; mHealth; smartphones; tablets; device usage; requirements; expectations; hospital; working equipment

Introduction

The usage of mobile devices, especially smartphones, has substantially increased, up to 95% in nearly all age groups in Germany [1]. With amazing advancement in mobile computer technology and connectivity, mobile devices have already revolutionized communication [2] as well as social media, mobility, fitness tracking, and further mobile health (mHealth) technologies, thus offering the potential to innovate health care inside and outside of clinical settings. The improvement of self-management as well as the effectiveness of the use of mHealth in professional medicine has been proven for several medical conditions, for example, the management of different chronic diseases such as arterial hypertension, diabetes or coronary heart disease, but also the management of acute diseases such as cardiac arrest and stroke [2-5]. Some authors even see smartphones as portable, multifunctional tools with the potential to become “the new stethoscope for physicians [6].” Increasingly, patients and the general population are being encouraged to take responsibility for their own health by actively monitoring their physiological parameters with smartphones, apps, and fitness trackers [7-10].

A fundamental component of the operationalization of mHealth is the usage of mobile devices, especially smartphones and tablets, by patients or health care professionals. The central areas of utilization of mobile devices in hospitals are (1) oral and written communication [11-15]; (2) documentation, organization, and information [16-19]; (3) decision support, notifications, and alarms [20-23], (4) education and professional training [24-28]; and (5) self-monitoring by physicians [29]. Taken together, physicians use mobile devices to assure their own decisions in a clinical environment and to increase efficiency in their workplaces [30].

The benefits of the usage of mobile devices in health care is counterbalanced by problems such as the colonization of surfaces with harmful pathogens or the distraction of medical staff [31,32]. Even more relevant are regulatory and organizational barriers for the implementation of mHealth apps on mobile devices in hospitals, missing standards for the development of health apps, information safety issues, and privacy concerns [12,33,34]. Furthermore, the parallel use of private devices for professional and private purposes is common [6], and thus further impacts data protection and patients’ privacy.

Both the combination use of mobile devices privately and on duty, and the physician’s attitude toward its deployment for so many different areas of utilization, make a structured systematic overview of actual needs difficult. There are hardly any data on the use of mobile devices in hospitals for clinical applications. Therefore, we aim to evaluate the current usage of mobile devices of physicians in German university medical centers and to explore their opinion and perceived needs regarding mobile devices.

Methods

Study Design, Data Collection, and Recruitment

For the preparation of the survey questions, an unstructured exploratory interview was conducted with 3 junior and 3 senior physicians focusing on their requirements and perceptions toward mobile devices. The results were used to construct the questionnaire for the actual study.

The study was designed as an open web-based survey in 8 German university hospitals and conducted among physicians of all medical disciplines (Limesurvey). To prove the functionality of the survey and the clarity of the questions, a test run was sent to a small group of 25 anesthesiologists and critical care physicians in June 2019. Minor remarks and improvement recommendations were made and integrated in the final version of the online survey which was sent via email with a link to the survey that was valid for 19 weeks. Responsible contact persons in the respective hospitals forwarded the link to physicians in their hospitals. During the period of data collection, no bug fixes and content changes were made, and no unexpected events such as system failures or server downtime were observed. The local data protection officer and the local ethics committee were consulted and had no concerns regarding the study. The title page of the survey contained information regarding the length, the foreseen time for completion, and the purpose the questionnaire. Completion of the survey was taken as consent for scientific usage of the collected data.

The survey was divided into sections, one of which contained questions about mobile devices usage. Biographical questions were included in another section (see Multimedia Appendix 1 for the English version of the survey).

Different types of survey questions were prepared: closed-ended questions, open-ended questions, rating questions, Likert scale questions (4-point scale), multiple choice questions, and demographic questions. Most questions allowed the participants to give multiple answers. Answering questions was not compulsory, as we expected that mandatory answers increases the risk that participants do not complete the web-based survey. Only fully completed questionnaires were included in the analysis. It was taken into account that some items were not answered by all participants (indicated as no response) resulting in a variation in the total number of answers. The survey instructions stated that cordless telephones (such as digital enhanced cordless telecommunications known as DECT) without additional functions and digital message receivers are not included in the survey to keep a narrower definition of a mobile device.

Statistical Analysis

As this study aimed to provide a general overview over physicians' attitudes and expectations toward mobile devices, data were predominantly analyzed with descriptive counts and proportions, applying significance tests only in a few selected cases. For nominal variables with a particularly large number of values (eg, medical discipline), values were summarized into broad categories, if possible. Data are given as absolute numbers or their percentages; summaries are given as median and as interquartile range for ordinal data and mean and standard deviation for continuous variables. Some of the survey items allowed for multiple responses (eg, choose all that apply), thus invalidating the use of classical chi-square testing to check for associations between those items. For significance testing of multiple response item associations, the nonparametric bootstrap variant of the simultaneous pairwise marginal independence test proposed by Bilder and Loughin [35] was used, which was implemented in the MRCV package (version 0.3-3) [36] in R (version 4.0.1). To test for associations of score variables with single-response items, the Kruskal-Wallis rank sum test was used.

The 33 medical disciplines of the participants were classified into 6 categories (see Multimedia Appendix 2, Table S1) to identify dependencies between the discipline categories and the survey answers. Furthermore, the mobile devices question group, which investigates the current usage, the needs, and requirements of physicians, was simplified. For the analysis, the tasks conducted with mobile devices in 5 fields of application in stationary hospital care (defined as all fields of inpatient care) were categorized as follows: (1) oral and written communication; (2) documentation, organization, and information desk; (3) decision support, notifications, and alarms; (4) Education and professional training; and (5) self-monitoring by physicians (see Table 1).

Table 1. Fields of application of mobile devices in stationary hospital care.

Categories	Functions
1. Oral/written communication	• Official phone calls • Official text messages (eg, SMS, messenger) • Web conferences (eg, tumor conferences)
2. Documentation, organization, and information desk	• Time scheduling and workflow support • Mobile EHR[a] to look up patient information and for medical documentation • Written instruction and recording procedures and examinations • Written inter/intraprofessional communication (doctors, nurses, therapists, consult requests) • To look up dosages, diagnoses and guidelines (online/offline)
3. Decision support, notifications, and alarms	• Alarming while monitoring of vital signs • (Early) warning system to prevent adverse effects (eg, pharmacological interaction) • Decision support and definition of therapies
4. Education and professional training	• Education system for job training, education and professional training
5. Self-monitoring by physicians	• Monitoring of own vital signs/motion analysis (eg, pedometer, energy consumption)

[a] EHR: electronic health record.

We developed a scoring system to analyze the participants' attitude toward mobile devices by assigning positive values to answers indicating a positive attitude toward mobile devices (fully disagree=0; fully agree=3) and negative values to EHR answers indicating a negative attitude (fully disagree=0; fully agree=–3). These values were summed for each participant and stratified by age groups, by medical disciplines, and by technical affinity.

Results

Demographic and Professional Characteristics

In total, 303 physicians with a mean clinical work experience of 12.7 years completed the survey. The full demographic and professional characteristics are given in Table 2. The participating physicians displayed a wide range of medical disciplines and the full range of discipline categories. The study population worked in all operational areas of the hospital, and physicians from all professional levels completed the web-based survey.

Table 2. Demographic and professional characteristics.

Characteristic	Value (n=303), n (%)
Age range (years)	
18-24	1 (0.3)
25-34	98 (32.3)
35-44	103 (34.0)
45-54	69 (22.8)
55-65	21 (6.9)
>65	3 (1.0)
No response	8 (2.6)
Gender	
Female	121 (39.9)
Male	173 (57.1)
No response	9 (3.0)
Current occupation	
Assistant physician	101 (33.3)
Medical specialist	49 (16.2)
Senior physician	108 (35.6)
Clinic director	28 (9.2)
Others	6 (2.0)
No response	11 (3.6)
Medical field/discipline	
Anesthesiology/intensive care medicine	75 (24.8)
Internal medicine	53 (17.5)
Pediatrics	25 (8.3)
Surgery	22 (7.3)
Neurology	14 (4.6)
Dermatology	12 (4.0)
Microbiology, virology, infectiology	10 (3.3)
Psychiatry and psychotherapy	10 (3.3)
Psychosomatic medicine and psychotherapy	8 (2.6)
Neurosurgery	8 (2.6)
Ophthalmology	7 (2.3)
Pathology	7 (2.3)
Otorhinolaryngology	5 (1.7)
Child and adolescent psychiatry and psychotherapy	5 (1.7)
Laboratory medicine	5 (1.7)
Radiology	5 (1.7)
Urology	5 (1.7)
Other disciplines/specialization	43 (14.2)
Predominant workplace	
Hospital ward	123 (40.6)
Operating theatre	106 (35.0)
Outpatient clinic	100 (33.0)

Characteristic	Value (n=303), n (%)
Intensive care unit	89 (29.4)
Office	50 (16.5)
Laboratory	33 (10.9)
Functional area	30 (9.9)
Others	15 (5.0)
Clinical professional experience (years), mean (SD)	12.7 (9.3)

Mobile Device Usage: Devices and Operation Purposes

Almost all respondents had smartphones (294/303, 97.0%) and laptops or desktops (280/303, 92.4%) for private use outside the working environment, 61.7% (187/303) of the respondents used tablets, and 20.8% (63/303) used wearables such as smartwatches and fitness trackers privately.

In clinical daily routine, 71.0% of physicians (215/303) used mobile devices. The operational purposes of mobile devices in clinical practice are widespread (Table 3). Predominantly, mobile devices are used for basic functions such as looking up information, oral and written communication (text messages, emails), and time scheduling.

Table 3. For which operation purposes do you use mobile devices in your clinical routine?

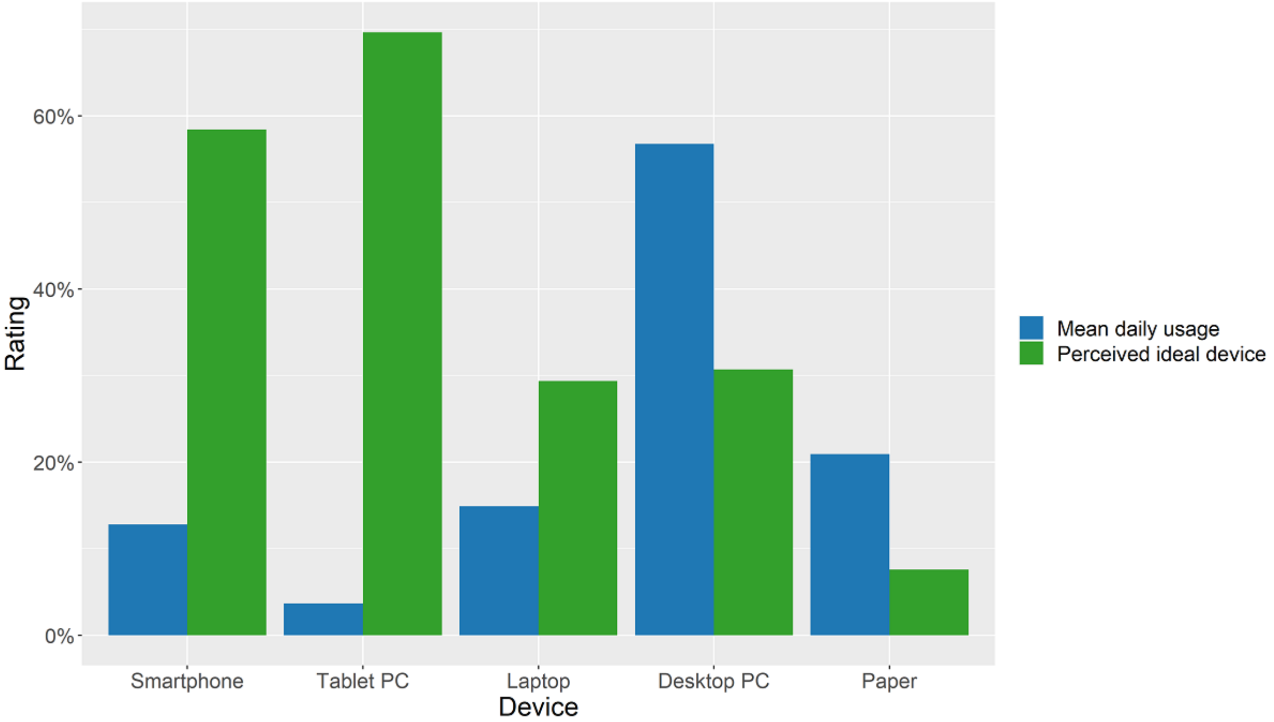
Operation purposes	Value (n=303), n (%)
Phone calls	154 (50.8)
Text messages (eg, SMS, messenger)	130 (42.9)
Email communication	157 (51.8)
Look up of information	190 (62.7)
Mobile access to hospital information systems	57 (18.8)
Time scheduling	141 (46.5)
Private communication	107 (35.3)
Dictation of texts	18 (5.9)
Scientific work	79 (26.1)
Other	14 (4.6)

Almost 79% of all respondents (238/303, 78.6%) stated that they used private devices for official uses in daily clinical routine. The majority of the respondents stated that they used their private mobile device because they are not provided with an official mobile device in hospital care (146/303, 48.2%). Other reasons were the ability to work at home (137/303, 45.2%) and the fact that mobile work is only possible with their private mobile devices (105/303, 34.7%). Some respondents also stated that their private devices are better than the official device (85/303, 28.1%) and that private communication is only possible with their private mobile device (65/303, 21.5%).

Perceived Ideal Device Versus Actual Time of Usage

The vast majority of participating physicians rated tablets (211/303, 69.6%) and smartphones (177/303, 58.4%) as the most appropriate device for their area of application (Figure 1). A smaller group of respondents (93/303, 30.7%) ranked desktop computers as most appropriate, and a group of respondents (89/303, 29.4%) regarded laptops as the most suitable device for their work area. Still 23 respondents (23/303, 7.6%) rated paper as most suitable for their professional work. Respondents could choose more than one answer for this question, and the simultaneous pairwise marginal independence test revealed no significant differences in answers between the different age groups (χ_s^2=34.19, *P*=.052).

Figure 1. Percentage of respondents rating device as ideal device versus mean percentage of daily worktime used.



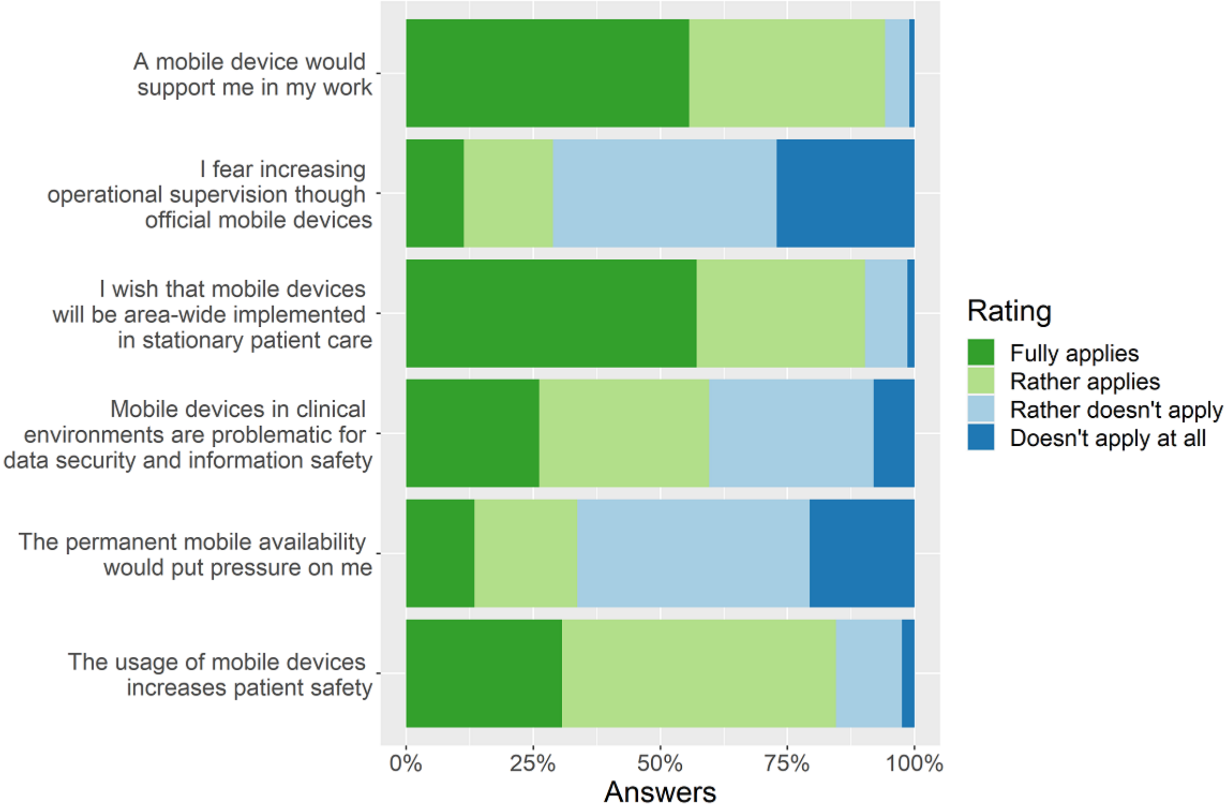
Contrasting their conception of ideal device usage, the physicians were asked to rate their actual time of usage of the devices during their daily total worktime. While most physicians rated smartphones and tablets as the perceived ideal device, in today’s clinical practice, desktop computers and paper are predominantly used (Figure 1).

Personal Opinion About Mobile Devices

In another survey section, we asked about the personal opinion of physicians toward mobile devices in stationary hospital care. We inquired whether mobile devices are supportive tools, whether they should be implemented in stationary hospital care and whether they increase patient safety. Furthermore, we asked for data security and information safety concerns regarding mobile devices and whether respondents fear an increasing workload or increasing operational supervision. For analysis, the statements “rather applies” and “fully applies” were taken as agreement. Most respondents (276/303, 91.1%) agreed with the statement “A mobile device would support me in my work.” Moreover, most physicians wish for the area-wide implementation of mobile devices in stationary patient care (259/303, 85.5%).

Regarding data security and information safety, a slight majority of 56.4% (171/303) of respondents expressed concerns. However, the large majority of 79.2% (240/303) of respondents agreed that the usage of mobile devices could increase patient safety (Figure 2).

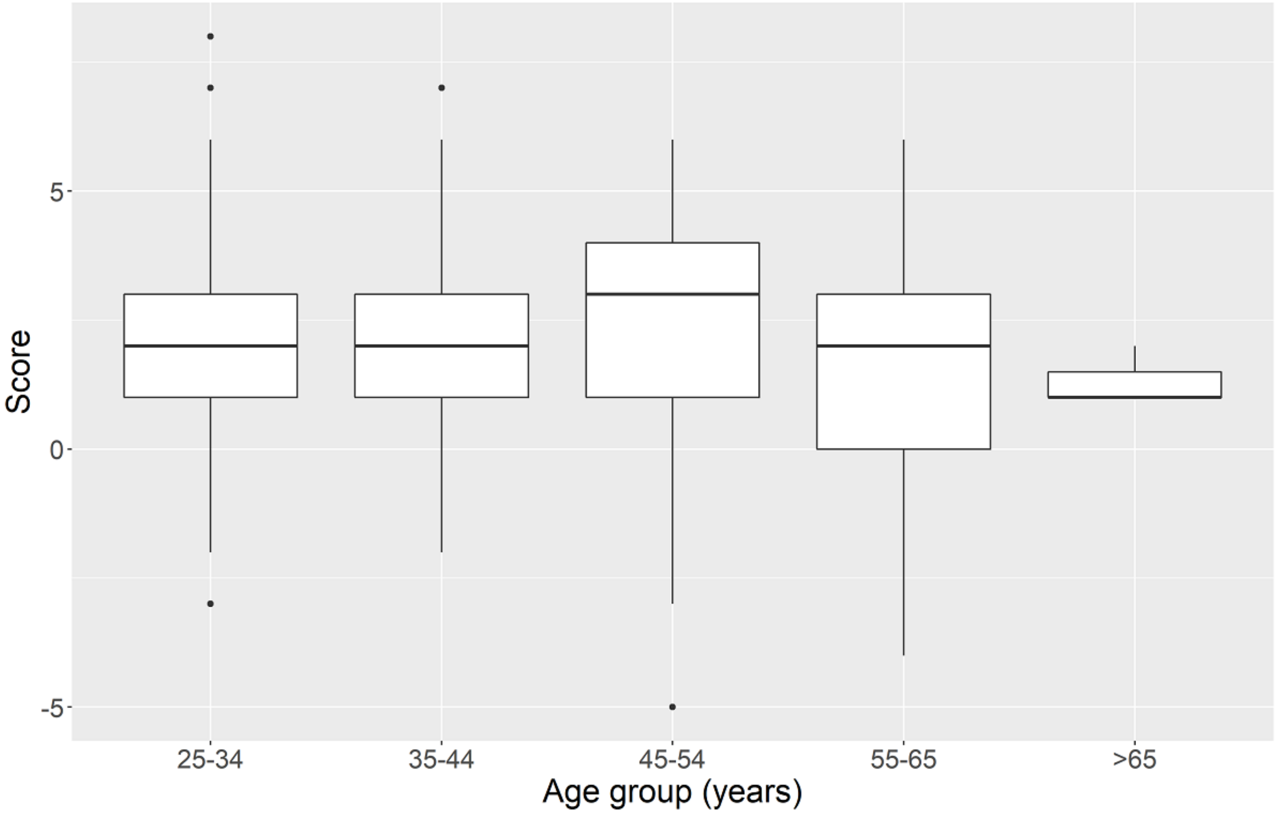
Figure 2. Personal opinion about mobile devices.



In sum, most physicians rated mobile devices as useful and supportive tools that should be implemented in stationary hospital care.

In general, a Kruskal-Wallis test did not show significant differences in the attitude toward mobile devices between age groups (H_5=7.29, P=.20; Figure 3).

Figure 3. Attitude toward mobile devices stratified by age groups.



Self-Reported Technical Affinity

Overall, the study population reported a high technical affinity level (mean 4 out of 5).

There was a significant association between technical affinity and the preference of device in medical care (χ_s^2=53.84, P<.001) showing that with increasing self-reported technical affinity the preference for smartphones and tablets increases compared to desktop computers (Figure 4).

The most optimistic respondents toward mobile devices were those who also had the highest self-reported technical affinity (Figure 5). With increasing technical affinity, the score for the attitude toward mobile devices increased (H_4=17.31, P=.002).

Figure 4. Device preference stratified by (self-reported) technical affinity level.

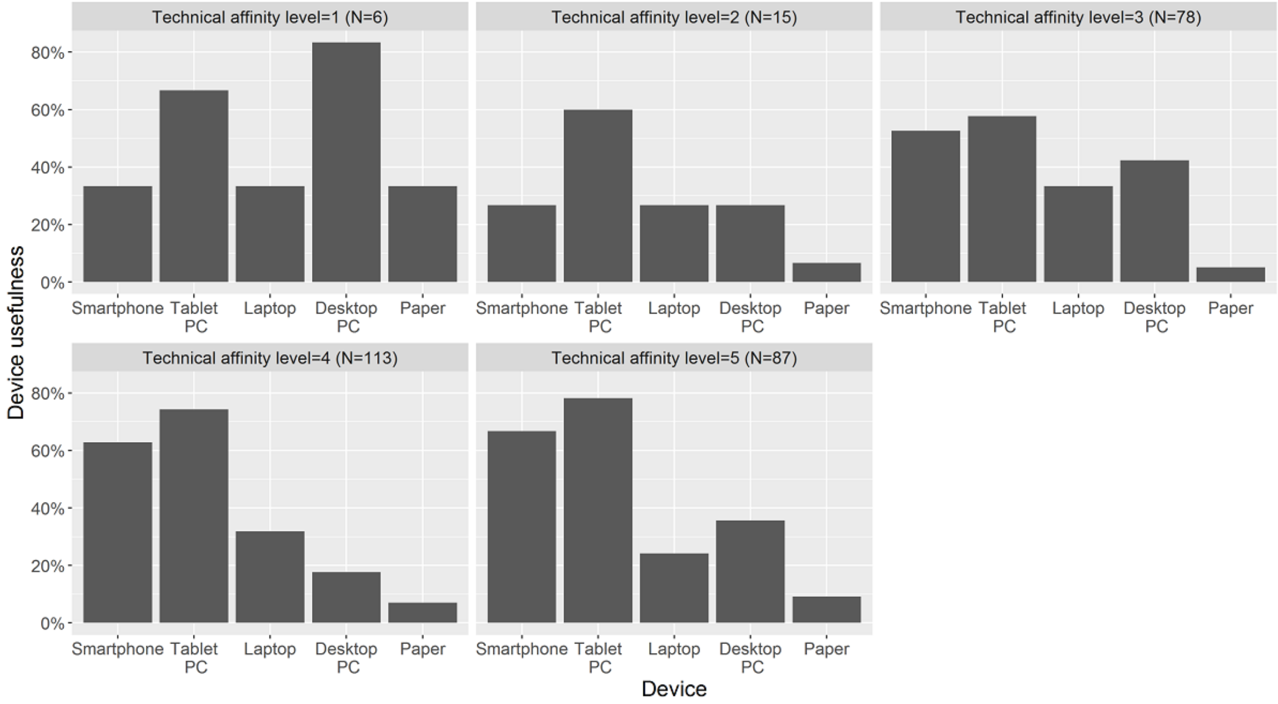
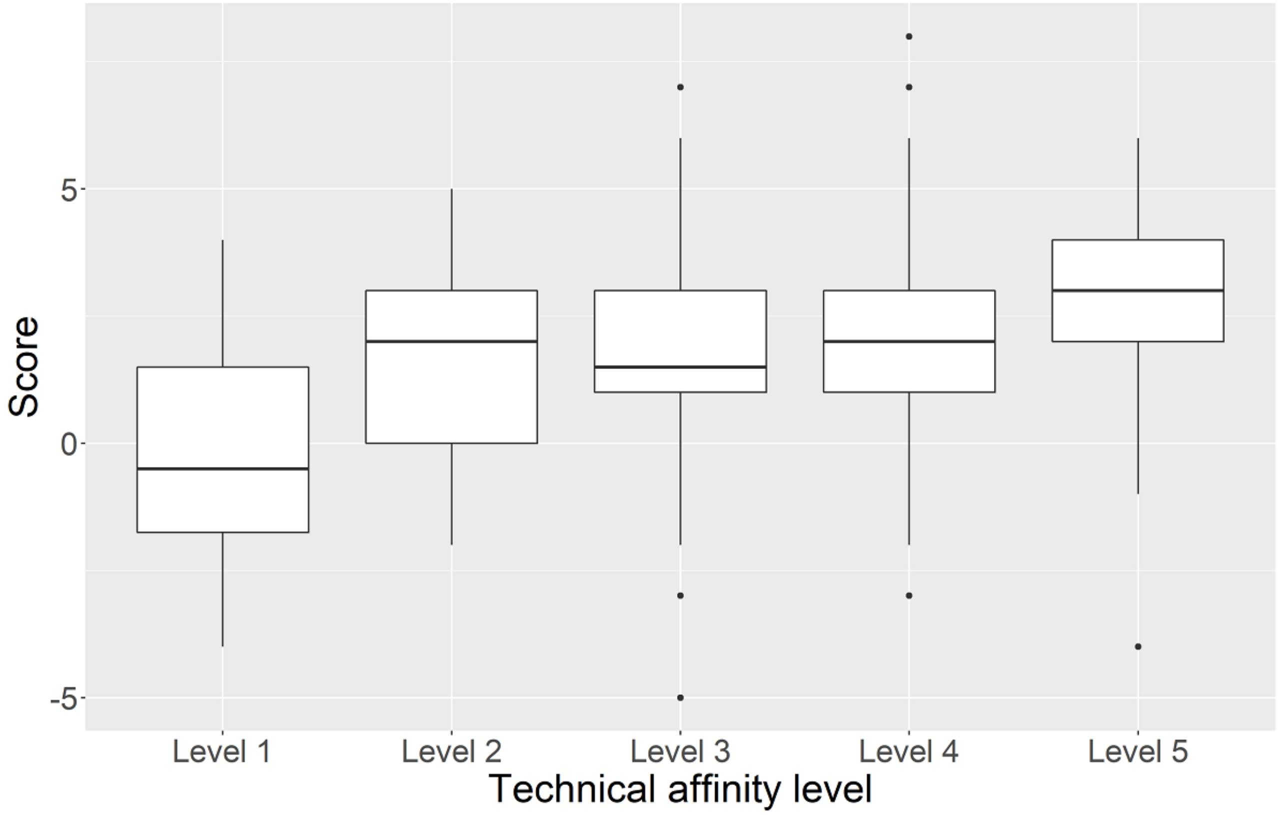


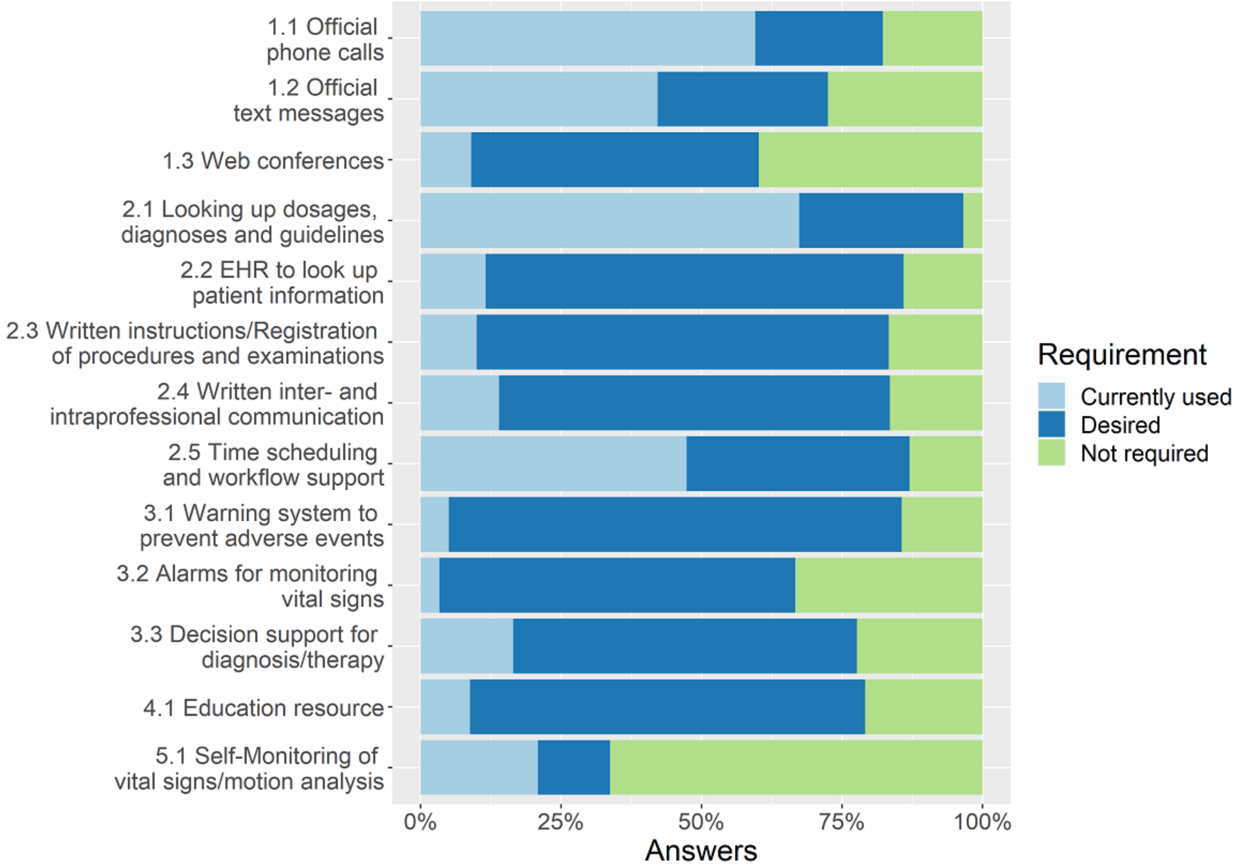
Figure 5. Attitude toward mobile devices stratified by (self-reported) technical affinity level.



Fields of Application and Desired Uses of Mobile Devices in Medicine

Finally, we examined the actual use of mobile devices and which functions the participants desired to use in stationary hospital care, if available. Five major fields of application of mobile devices could be identified (Table 1). Within these fields, the following 4 functions were the most commonly used mobile device apps by physicians: official phone calls (171/303, 56.4%); official text messages (118/303, 38.9%); looking up dosages, diagnoses, and guidelines (194/303, 64.0%); and time scheduling and workflow support (135/303, 44.6%). The most desired mobile device uses were for (early) warning systems to prevent adverse effects in hospital care (eg, pharmacological interactions; 224/303, 73.9%), for mobile EHR to look up patient information and for medical documentation (211/303, 69.6%), and for written instructions and recording procedures and examinations (206/303, 68.0%). The results can be found in Figure 6.

Figure 6. Fields of application and desired functions. EHR: electronic health record.



Discussion

Current Situation and Physicians' Requirements

In today's clinical practice, physicians are confronted with increasing amounts of patient information, information overload and information inaccessibility, endangering patient safety and potentially leading to fatal errors [37]. A substantial proportion of physicians use their private mobile devices for professional purposes, especially because they are not provided with official mobile devices for their work. Our survey results show that physicians expect that an official mobile device would support them in their work and increase patient safety. Most participants used mobile devices in hospitals for communication and organization (phone calls, text messages, time scheduling, and information), and thus, to increase the efficiency in their workplaces [30]. Among our study population, most physicians would prefer to use official devices but accept using their private mobile devices for professional purposes in stationary hospital care if there is no alternative.

Potential Benefits for Physicians and Patient Care

Currently, physicians in hospital care spend a large amount of their worktime performing documentation in clinical information systems [38,39]. Therefore, physicians want to use mobile devices for documentation and recording procedures in these systems. Today, the tasks of documentation and recording procedures are most commonly performed with a desktop computer and cannot be performed in equal quality on a smaller mobile device. Thus, not every task in clinical care can be transferred from a desktop computer to a smartphone or a tablet without changing the process. Drews et al [16] examined the impact of the form factor of various mobile devices and desktop computers on the usability of EHRs; the authors concluded that even the largest form of a mobile device does not perform as well as a desktop computer for the usage of EHR. Consequently, the processes of documentation and recording procedures need to be changed before they can be performed in equal quality on a tablet or a smartphone in hospital care.

Smartphones and tablets at the point of care could support in-hospital physicians (eg, with functionalities such as taking notes via voice recognition for clinical documentation instead of keyboard-and-mouse interface) [40,41]. Payne et al [42] described the implementation of a smartphone-based system as a supporting system with automated speech recognition integrated in a commercial EHR; the mobile supporting system that was described has the potential to reduce the documentation burden of doctors and nurses, which is perceived as one of the big problems in today's health care [38,39]. We assume that physicians would benefit from mobile devices with automated speech recognition for documentation in and access to clinical information systems. Nevertheless, 48.2% (146/303) of all respondents are not provided with official mobile devices for these uses in hospital care and, as shown in Figure 6, less than 11% are already using a mobile EHR in our study population. As a precondition of continual documentation of patient information, mobile apps need to be integrated in central information systems in hospitals by using technical and semantic interoperability standards.

German hospitals are no exception in the rare use of mobile devices even for basic tasks, such as documentation in clinical information systems. However, physicians are open to an increasing use of more advanced mobile uses such as decision support systems on mobile devices in their clinical practice. Large proportions of physicians stated the wish to be able to use (early) warning systems for prevention of adverse effects, possibly harmful pharmacological interactions, smart monitoring of vital signs and decision support, and definition of therapies in clinical practice.

Regulatory Challenges

The development of medical software such as mobile apps, especially those subject to new European Medical Device Regulation 2017/745, is a complex process including many obligations and requirements for manufacturers [43]. In addition to the regulatory legal requirements of the European Medical Device Regulation, health care professionals ask for the certification of health care apps [44].

The majority of physicians are aware that mobile devices may have implications on data security and information safety, while at the same time increasing patient safety. To ensure data security and information safety in hospitals, we suggest that apps that support health care professionals in performing complex and critical tasks should be installed and operated on official smartphones and tablets. For fulfilling information security standards, the devices should be provided and administrated through the hospital information technology departments. This could also help to reduce the potential legal grey area of using private mobile devices in the clinical environments for official purposes. Therefore, hospitals should implement a mobile device management process to safeguard the secure operation of mobile devices. Due to the increasing complexity of mobile device uses and increasing competencies of physicians in using mHealth apps, there is a need for a process to implement, teach, supervise, and evaluate clinical mHealth as well as mobile device and app competencies [24,45]. Finally, the added value of apps on mobile devices for physicians should be scientifically proven before being implemented in hospital care.

Strengths and Limitations

This web-based survey covered the use of mobile devices for physicians in stationary hospital care in 8 university hospitals in different regions in Germany. As far as we know, it is the first survey covering this study population and evaluating the usage of, requirements for, and expectations toward mobile devices. University hospitals accommodate all medical disciplines and physicians involved in patient care, research, teaching, and training. Thus, a wide range of medical disciplines was covered by the respondents.

Usage of and requirements for mobile devices such as mobile (telemonitoring) apps for patients (self-monitoring and mobile self-reporting [46]) or diagnostic instruments connected to a mobile device (eg, iECG or handheld ultrasound [3]) are not the subject matter of this study. We also did not distinguish between official mobile devices as personalized or shared mobile devices.

A limitation of the recruitment method of the web-based survey is potential volunteer bias. The mean self-reported technical affinity of the study population was 4 out of 5, suggesting that most participating physicians had a relatively high affinity for technology, in general and in medical practice. Consequently, further research, using paper-based and personal-oral survey methods, is needed to reach physicians with a lower technical affinity.

Conclusions

So far, the widespread use of official mobile devices such as smartphones and tablets has not become reality in stationary hospital care in German (university) hospitals. As long as physicians are predominantly using their private devices in clinical care, the usage of advanced apps with a deeper integration into the clinical information system infrastructure to support physicians remains uncertain.

Nevertheless, among the participating physicians of German university hospitals, technical affinity is high, and they have a very positive attitude toward mobile devices for clinical care. With our results, we demonstrated that the majority of the participating physicians used mobile devices for basic functionalities in hospital care. Although most physicians would prefer to work with mobile devices for documentation, writing instructions, and recording procedures in clinical information systems, a desktop computer is generally used for these tasks. Furthermore, physicians are willing to use their mobile devices for more progressive uses such as decision support systems or early warning systems. Thus, reasons for the low usage of official mobile devices in German hospitals are not the potential users, but rather regulatory, financial and organizational challenges, and missing interoperability standards.

While most physicians think that mobile devices would support their work and increase patient safety, they also mentioned concerns regarding data security and information safety.

Acknowledgments

This publication of the Smart Medical Information Technology for Healthcare consortium was supported by the German Federal Ministry of Education and Research (grant numbers 01ZZ1803B and 01ZZ1803C).

Authors' Contributions

OM and SF designed the survey and set up the web-based survey. OM, JG, and SF analyzed the data. OM wrote the manuscript. SF, JB, JK, JG, SD, and GM revised the article. All authors approved the final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey on the usage of mobile devices in stationary hospital care.
[DOC File , 113 KB - jmir_v22i12e23955_app1.doc]

Multimedia Appendix 2
Categories of medical disciplines.
[DOC File , 48 KB - jmir_v22i12e23955_app2.doc]

References

1. Tenzer F. Anzahl der Nutzer von Smartphones in Deutschland bis 2019. Statista. 2019. URL: https://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenutzer-in-deutschland-seit-2010/ [accessed 2020-07-20]
2. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
3. Bhavnani SP, Narula J, Sengupta PP. Mobile technology and the digitization of healthcare. Eur Heart J 2016 May 07;37(18):1428-1438 [FREE Full text] [doi: 10.1093/eurheartj/ehv770] [Medline: 26873093]
4. Triantafyllidis A, Kondylakis H, Votis K, Tzovaras D, Maglaveras N, Rahimi K. Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: A review of systematic reviews. Int J Med Inform 2019 Dec;132:103984. [doi: 10.1016/j.ijmedinf.2019.103984] [Medline: 31605884]
5. Sarfo FS, Ovbiagele B. Mobile health for stroke: a promising concept for research and practice. Mhealth 2017;3:4 [FREE Full text] [doi: 10.21037/mhealth.2017.02.01] [Medline: 28300225]
6. Dimond R, Bullock A, Lovatt J, Stacey M. Mobile learning devices in the workplace: 'as much a part of the junior doctors' kit as a stethoscope'? BMC Med Educ 2016 Aug 17;16(1):207 [FREE Full text] [doi: 10.1186/s12909-016-0732-z] [Medline: 27530343]
7. Vandelanotte C, Van Itallie A, Brown W, Mummery WK, Duncan MJ. Every step counts: understanding the success of implementing the 10,000 steps project. Stud Health Technol Inform 2020 Mar 02;268:15-30. [doi: 10.3233/SHTI200003] [Medline: 32141876]
8. Ahuja N, Ozdalga E, Aaronson A. Integrating mobile fitness trackers into the practice of medicine. Am J Lifestyle Med 2017;11(1):77-79 [FREE Full text] [doi: 10.1177/1559827615583643] [Medline: 30202316]
9. Saxon LA. Mobile health application solutions. Circ Arrhythm Electrophysiol 2016 Feb;9(2):e002477. [doi: 10.1161/CIRCEP.115.002477] [Medline: 26810595]
10. Goh G, Tan NC, Malhotra R, Padmanabhan U, Barbier S, Allen JC, et al. Short-term trajectories of use of a caloric-monitoring mobile phone app among patients with type 2 diabetes mellitus in a primary care setting. J Med Internet Res 2015 Feb 03;17(2):e33 [FREE Full text] [doi: 10.2196/jmir.3938] [Medline: 25648130]
11. Liu X, Sutton PR, McKenna R, Sinanan MN, Fellner BJ, Leu MG, et al. Evaluation of secure messaging applications for a health care system: a case study. Appl Clin Inform 2019 Jan;10(1):140-150 [FREE Full text] [doi: 10.1055/s-0039-1678607] [Medline: 30812040]
12. Rokadiya S, McCaul JA, Mitchell DA, Brennan PA. Leading article: Use of smartphones to pass on information about patients - what are the current issues? Br J Oral Maxillofac Surg 2016 Jul;54(6):596-599. [doi: 10.1016/j.bjoms.2016.04.020] [Medline: 27185230]
13. Househ M. The role of short messaging service in supporting the delivery of healthcare: an umbrella systematic review. Health Informatics J 2016 Jun;22(2):140-150 [FREE Full text] [doi: 10.1177/1460458214540908] [Medline: 25038203]
14. Mars M, Scott RE. WhatsApp in clinical practice: a literature review. Stud Health Technol Inform 2016;231:82-90. [Medline: 27782019]

15. Wu R, Appel L, Morra D, Lo V, Kitto S, Quan S. Short message service or disService: issues with text messaging in a complex medical environment. Int J Med Inform 2014 Apr;83(4):278-284. [doi: 10.1016/j.ijmedinf.2014.01.003] [Medline: 24495802]
16. Drews FA, Zadra JR, Gleed J. Electronic health record on the go: device form factor and Fitts' law. Int J Med Inform 2018 Mar;111:37-44. [doi: 10.1016/j.ijmedinf.2017.12.010] [Medline: 29425632]
17. Saleem JJ, Savoy A, Etherton G, Herout J. Investigating the need for clinicians to use tablet computers with a newly envisioned electronic health record. Int J Med Inform 2018 Feb;110:25-30. [doi: 10.1016/j.ijmedinf.2017.11.013] [Medline: 29331252]
18. Schooley B, Walczak S, Hikmet N, Patel N. Impacts of mobile tablet computing on provider productivity, communications, and the process of care. Int J Med Inform 2016 Apr;88:62-70. [doi: 10.1016/j.ijmedinf.2016.01.010] [Medline: 26878764]
19. Motulsky A, Wong J, Cordeau J, Pomalaza J, Barkun J, Tamblyn R. Using mobile devices for inpatient rounding and handoffs: an innovative application developed and rapidly adopted by clinicians in a pediatric hospital. J Am Med Inform Assoc 2017 Apr 01;24(e1):e69-e78 [FREE Full text] [doi: 10.1093/jamia/ocw107] [Medline: 27554824]
20. Richardson KM, Fouquet SD, Kerns E, McCulloh RJ. Impact of mobile device-based clinical decision support tool on guideline adherence and mental workload. Acad Pediatr 2019;19(7):828-834 [FREE Full text] [doi: 10.1016/j.acap.2019.03.001] [Medline: 30853573]
21. Machado JP, Lam XT, Chen J. Use of a clinical decision support tool for the management of traumatic dental injuries in the primary dentition by novice and expert clinicians. Dent Traumatol 2018 Apr;34(2):120-128. [doi: 10.1111/edt.12390] [Medline: 29476702]
22. Huang CY, Cheng PL, Chang PL. Using a mobile device application to support emergency clinicians in diagnosing pulmonary embolism. Stud Health Technol Inform 2017;245:1363. [Medline: 29295442]
23. Dexheimer JW, Borycki EM. Use of mobile devices in the emergency department: a scoping review. Health Informatics J 2015 Dec;21(4):306-315 [FREE Full text] [doi: 10.1177/1460458214530137] [Medline: 24782479]
24. Hilty DM, Chan S, Torous J, Luo J, Boland RJ. Mobile health, smartphone/device, and apps for psychiatry and medicine: competencies, training, and faculty development issues. Psychiatr Clin North Am 2019 Sep;42(3):513-534. [doi: 10.1016/j.psc.2019.05.007] [Medline: 31358129]
25. Mackay BJ, Anderson J, Harding T. Mobile technology in clinical teaching. Nurse Educ Pract 2017 Jan;22:1-6. [doi: 10.1016/j.nepr.2016.11.001] [Medline: 27871040]
26. Han H, Nelson E, Wetter N. Medical students' online learning technology needs. Clin Teach 2014 Feb;11(1):15-19. [doi: 10.1111/tct.12092] [Medline: 24405913]
27. Bedi HS, Yucel EK. "I just bought my residents iPads… now what?" the integration of mobile devices into radiology resident education. AJR Am J Roentgenol 2013 Oct;201(4):704-709. [doi: 10.2214/AJR.13.10674] [Medline: 24059358]
28. Tanaka PP, Hawrylyshyn KA, Macario A. Use of tablet (iPad®) as a tool for teaching anesthesiology in an orthopedic rotation. Rev Bras Anestesiol 2012;62(2):214-222 [FREE Full text] [doi: 10.1016/S0034-7094(12)70119-8] [Medline: 22440376]
29. DiFrancisco-Donoghue J, Jung M, Stangle A, Werner WG, Zwibel H, Happel P, et al. Utilizing wearable technology to increase physical activity in future physicians: a randomized trial. Prev Med Rep 2018 Dec;12:122-127 [FREE Full text] [doi: 10.1016/j.pmedr.2018.09.004] [Medline: 30234000]
30. Nerminathan A, Harrison A, Phelps M, Alexander S, Scott KM. Doctors' use of mobile devices in the clinical setting: a mixed methods study. Intern Med J 2017 Mar;47(3):291-298. [doi: 10.1111/imj.13349] [Medline: 27925381]
31. Ulger F, Esen S, Dilek A, Yanik K, Gunaydin M, Leblebicioglu H. Are we aware how contaminated our mobile phones with nosocomial pathogens? Ann Clin Microbiol Antimicrob 2009 Mar 06;8:7 [FREE Full text] [doi: 10.1186/1476-0711-8-7] [Medline: 19267892]
32. Attri JP, Khetarpal R, Chatrath V, Kaur J. Concerns about usage of smartphones in operating room and critical care scenario. Saudi J Anaesth 2016;10(1):87-94 [FREE Full text] [doi: 10.4103/1658-354X.169483] [Medline: 26952181]
33. van Velthoven MH, Wyatt JC, Meinert E, Brindley D, Wells G. How standards and user involvement can improve app quality: a lifecycle approach. Int J Med Inform 2018 Oct;118:54-57. [doi: 10.1016/j.ijmedinf.2018.08.003] [Medline: 30153922]
34. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
35. Bilder CR, Loughin TM. Testing for marginal independence between two categorical variables with multiple responses. Biometrics 2004 Mar;60(1):241-248. [doi: 10.1111/j.0006-341X.2004.00147.x] [Medline: 15032795]
36. Koziol NA, Bilder CR. MRCV: Methods for Analyzing Multiple Response Categorical Variables (MRCVs). R Package Version 0.3-3. 2014. URL: https://CRAN.R-project.org/package=MRCV [accessed 2020-07-20]
37. Pickering BW, Gajic O, Ahmed A, Herasevich V, Keegan MT. Data utilization for medical decision making at the time of patient admission to ICU. Crit Care Med 2013 Jun;41(6):1502-1510. [doi: 10.1097/CCM.0b013e318287f0c0] [Medline: 23528804]
38. Tawfik DS, Profit J, Webber S, Shanafelt TD. Organizational factors affecting physician well-being. Curr Treat Options Pediatr 2019 Mar;5(1):11-25 [FREE Full text] [doi: 10.1007/s40746-019-00147-6] [Medline: 31632895]

39. Collins S, Couture B, Kang MJ, Dykes P, Schnock K, Knaplund C, et al. Quantifying and visualizing nursing flowsheet documentation burden in acute and critical care. AMIA Annu Symp Proc 2018;2018:348-357 [FREE Full text] [Medline: 30815074]
40. Lorenzetti DL, Quan H, Lucyk K, Cunningham C, Hennessy D, Jiang J, et al. Strategies for improving physician documentation in the emergency department: a systematic review. BMC Emerg Med 2018 Oct 25;18(1):36 [FREE Full text] [doi: 10.1186/s12873-018-0188-z] [Medline: 30558573]
41. Collier R. Rethinking EHR interfaces to reduce click fatigue and physician burnout. CMAJ 2018 Aug 20;190(33):E994-E995 [FREE Full text] [doi: 10.1503/cmaj.109-5644] [Medline: 30127043]
42. Payne TH, Alonso WD, Markiel JA, Lybarger K, White AA. Using voice to create hospital progress notes: Description of a mobile application and supporting system integrated with a commercial electronic health record. J Biomed Inform 2018 Jan;77:91-96 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.004] [Medline: 29233669]
43. Keutzer L, Simonsson US. Medical device apps: an introduction to regulatory affairs for developers. JMIR Mhealth Uhealth 2020 Jun 26;8(6):e17567 [FREE Full text] [doi: 10.2196/17567] [Medline: 32589154]
44. Mayer MA, Rodríguez Blanco O, Torrejon A. Use of health apps by nurses for professional purposes: web-based survey study. JMIR Mhealth Uhealth 2019 Nov 01;7(11):e15195 [FREE Full text] [doi: 10.2196/15195] [Medline: 31682587]
45. Hilty D, Chan S, Torous J, Luo J, Boland R. A framework for competencies for the use of mobile technologies in psychiatry and medicine: scoping review. JMIR Mhealth Uhealth 2020 Feb 21;8(2):e12229 [FREE Full text] [doi: 10.2196/12229] [Medline: 32130153]
46. Triantafyllidis AK, Velardo C, Salvi D, Shah SA, Koutkias VG, Tarassenko L. A survey of mobile phone sensing, self-reporting, and social sharing for pervasive healthcare. IEEE J Biomed Health Inform 2017 Jan;21(1):218-227. [doi: 10.1109/JBHI.2015.2483902] [Medline: 26441432]

Abbreviations

EHR: electronic health record
mHealth: mobile health

Edited by G Eysenbach, R Kukafka; submitted 03.09.20; peer-reviewed by L van Velsen; comments to author 25.09.20; revised version received 06.10.20; accepted 28.10.20; published 21.12.20.

Please cite as:
Maassen O, Fritsch S, Gantner J, Deffge S, Kunze J, Marx G, Bickenbach J
Future Mobile Device Usage, Requirements, and Expectations of Physicians in German University Hospitals: Web-Based Survey
J Med Internet Res 2020;22(12):e23955
URL: http://www.jmir.org/2020/12/e23955/
doi:10.2196/23955
PMID:33346735

Original Paper

Prevalence of Mental Illnesses in Domestic Violence Police Records: Text Mining Study

George Karystianis[1], BSc, MSc, PhD; Annabeth Simpson[2], Bsc (Hons), MPH; Armita Adily[1], BMedSc, MPH, PhD; Peter Schofield[3], BSc (Hons), BS, Med, MSc, MD; David Greenberg[4], MB, ChB, FFPsych (SA), MMed, FRCPC, FRANZCP; Handan Wand[5], BSc, MSc, PhD; Goran Nenadic[6], BSc, MSc, PhD; Tony Butler[1], BSc, MSc, PhD

[1]School of Population Health, University of New South Wales, Sydney, Australia
[2]National Drug and Alcohol Research Centre, Sydney, Australia
[3]Neuropsychiatry Service, Hunter New England Health, Newcastle, Australia
[4]School of Psychiatry, University of New South Wales, Sydney, Australia
[5]Kirby Institute, University of New South Wales, Sydney, Australia
[6]School of Computer Science, University of Manchester, Manchester, United Kingdom

Corresponding Author:
George Karystianis, BSc, MSc, PhD
School of Population Health
University of New South Wales
Level 3, Samuels Building, Gate 11
Botany Street, UNSW Kensington Campus
Sydney, 2052
Australia
Phone: 61 (2) 9385 3811
Email: g.karystianis@unsw.edu.au

Abstract

Background: The New South Wales Police Force (NSWPF) records details of significant numbers of domestic violence (DV) events they attend each year as both structured quantitative data and unstructured free text. Accessing information contained in the free text such as the victim's and persons of interest (POI's) mental health status could be useful in the better management of DV events attended by the police and thus improve health, justice, and social outcomes.

Objective: The aim of this study is to present the prevalence of extracted mental illness mentions for POIs and victims in police-recorded DV events.

Methods: We applied a knowledge-driven text mining method to recognize mental illness mentions for victims and POIs from police-recorded DV events.

Results: In 416,441 police-recorded DV events with single POIs and single victims, we identified 64,587 events (15.51%) with at least one mental illness mention versus 4295 (1.03%) recorded in the structured fixed fields. Two-thirds (67,582/85,880, 78.69%) of mental illnesses were associated with POIs versus 21.30% (18,298/85,880) with victims; depression was the most common condition in both victims (2822/12,589, 22.42%) and POIs (7496/39,269, 19.01%). Mental illnesses were most common among POIs aged 0-14 years (623/1612, 38.65%) and in victims aged over 65 years (1227/22,873, 5.36%).

Conclusions: A wealth of mental illness information exists within police-recorded DV events that can be extracted using text mining. The results showed mood-related illnesses were the most common in both victims and POIs. Further investigation is required to determine the reliability of the mental illness mentions against sources of diagnostic information.

(J Med Internet Res 2020;22(12):e23725) doi:10.2196/23725

KEYWORDS

text mining; mental illnesses; domestic violence; police data; trend analysis

Introduction

Domestic violence (DV) is defined as “any incident of threatening behavior, violence or (psychological, physical, sexual, financial, emotional) abuse between adults who are or have been an intimate partner or family member, regardless of gender or sexuality” [1]. It can also occur in other relationships such as between caregivers and a dependent person (and vice versa) or those living together in a household [2]. According to the World Health Organization’s multicountry study of violence, the prevalence of physical and sexual partner violence toward women ranges from 15% to 71% globally [3,4]. In Australia, in 2018, 1 out of 6 women and 1 out of 16 men experienced physical or sexual violence or both by a current or previous partner [5] and on average, 1 woman a week is murdered by her current/former partner [6]. In addition, research has shown that children exposed to DV experience long-term effects on their development with increased risk of mental health issues, learning difficulties, and behavioral problems [7]. DV puts significant economic and health burden on the community and its prevention should be a public health priority [8]. Estimates have suggested that the annual financial burden in Australia arising from DV against women and their children is over AUD 22 billion (~US $16.3 billion), £66 billion (~US $89 billion) in the United Kingdom, and US $55 billion in the United States [9-11].

DV has been linked to significant comorbidity and mortality with both short- and long-term health consequences, particularly among women [3,8,12]. Evidence has shown that any immediate injury or trauma suffered in a DV setting has longer-term negative effects on the survivor’s well-being, contributing to poor health outcomes including post-traumatic stress disorder, chronic substance use, risky sexual behaviors, eating disorders, suicidal tendencies and attempts, as well as exacerbation of psychotic symptoms [3,4,12].

Associations have been found between mental health conditions (eg, bipolar disorder, schizophrenia) and the perpetration of violence toward others fueling perceptions that label these individuals as *dangerous*, leading to the stigmatization of this group [13-19]. Increasing evidence suggests that people with mental illness and psychiatric symptoms, however, are at a greater risk of victimization when compared to those without such symptoms [3,8,12,13,20-22]. Women with disabilities, including those with chronic mental or emotional conditions, experience higher rates of violent victimization than men with disabilities and women in the general population [23]. Men and women with severe mental illness (such as psychotic disorders) are two to eight times more likely to experience any form of DV abuse and to suffer poor health outcomes (eg, suicide attempt, substance abuse) than the general population [20,24]. This suggests the potential importance of knowing whether an individual has a pre-existing mental illness at the time of a DV event to enable prevention and intervention measures to be enacted.

The New South Wales Police Force (NSWPF) attends and subsequently records thousands of DV events each year—123,330 such events were recorded in 2017—in free text in their WebCOPS database, an online interface for the Computerized Operational Policing System (COPS) that enables the police to capture and analyze crime information on an organization-wide basis (NSWPF, personal communication). These police-recorded DV events contain a wealth of unutilized mental illness information for persons of interest (POIs)—individuals involved in a DV event that have been accused or charged for perpetrating DV related crimes—and victims that could be used to identify trends in those involved in DV and assist in shaping early DV intervention and prevention policies. However, the vast number of such events make the manual extraction of potentially useful information with traditional ethnographic/qualitative approaches impractical. Indeed, one recent research paper commented that “…there is no systematic way to extract information from these [police] narratives other than by manual review” [25].

Automated methods for large-scale processing of free text known as text mining have been used for over 30 years to harvest information from unstructured text in many domains, including medicine [26,27]. Several attempts have been made to extract mental health–related information from various free-text resources [28-33] including identification of drug side effects from psychiatric narratives by applying rule- and dictionary-based methods and machine learning approaches [28,30,31]. There have been efforts to extract treatment outcomes for major depressive disorders from electronic medical records with a supervised approach combined with logistic regression [29], whereas Jackson et al [33] and Karystianis et al [32] identified psychiatric symptoms from clinical discharge summaries and psychiatric records using regular expression pattern matching and a rule-based approach, respectively. Most recently, Wu et al [34] applied dictionary and machine learning methods to extract depressive symptoms in order to validate the diagnosis of major depressive disorders from electronic health records.

Because of the rapid implementation of automated technologies in various fields, text mining has been identified as a potential tool of interest in the analysis of police data. However, there are relatively few text mining methods that have been developed to analyze police narratives. Recent work has been conducted in automatically processing police reports to identify information of interest [35-38]. Attempts have been made to automatically identify offenders’ names, illicit drugs, and weapons with various degrees of success from police narrative reports through named entity extractors [35,36], while others aimed to classify police reports as DV or non-DV related using an unsupervised clustering method [37]. Most recently, deep learning methods have been used to extract mental health–related incidents from police narratives with an 89% accuracy [38].

This study builds on our previously published work that focused solely on the design, description, and evaluation of the text mining methodology [39]. We present the extracted mental illness mentions from 416,441 police-recorded DV events that involve single POIs and victims by age groups and sex, and compare the prevalence of the identified information with a fixed field mental illness flag also recorded in the WebCOPS system for the same cohort, as well as with the national estimated prevalence of mental illness in Australia. To the best of our knowledge, this is the first attempt to report automatically extracted mental health information from a large cohort of police-recorded DV events.

Methods

Data

Information relating to DV events that the police attend is recorded in their WebCOPS database as both structured data (fixed fields) covering demographic information (eg, name, date of birth, Aboriginal status, whether weapons were used) and free-text *event narratives*. Each police-recorded DV event contains at least one event narrative which details the incident(s) between the POI and victim, covering the circumstances of the event, whether alcohol, drugs, or both were involved, and any action(s) taken by the police. The text narratives can contain misspellings and typographical errors, often with informal acronyms, jargon, and abbreviations, that may bear ambiguous meanings depending on the context. Typically, they are used as an aide-memoire for the police and lawyers should the case proceed to court and by lawyers in court proceedings, but they have not been utilized in a substantive manner for research purposes.

We obtained 492,393 police-recorded DV events from the NSWPF from January 2005 to December 2016 that were flagged in the fixed fields with one of the following tags: "domestic" as the type of offence, "domestic violence related" as the associated factor of the police event, or the relationship status between the victim and the POI being described as spouse/partner (including ex-spouse/ex-partner), boy/girlfriend (including ex-boy/ex-girlfriend), parent/guardian (including step/foster), child (including step/foster), sibling, other member of family (including kin), or carer. These police-recorded DV events covered the following incident categories: assaults, breaches of Apprehended Violence Orders, homicides, malicious damage to property, and offences against another person such as intimidation, kidnapping, abduction, or harassment. The police-recorded DV events also contained incidents where no crime was committed but the police did attend and record the event. A hypothetical deidentified police-recorded DV event is shown in Multimedia Appendix 1.

Ethics

Permission to access the police-recorded DV events was granted by the NSWPF following ethics approval from the University of NSW Human Research Ethics Committee (HC16558).

Extraction and Normalization of Mental Illness Mentions

We designed and applied a text mining methodology that was implemented through the General Architecture for Text Engineering (GATE; 8.4.1 version) [40], a text mining framework to capture mental illness mentions (including traumatic brain injury and dementia) for POIs and victims. GATE was selected because it supports the development of rule-based approaches as it can easily manipulate unstructured data. We developed 2 sets of rules: based on common lexical patterns observed in the text of 200 police-recorded DV events that indicate the presence of a specific mental health mention for a POI (eg, "POI is suffering from dementia") or for a victim (eg, "the victim was diagnosed with paranoid schizophrenia") and based on related semantic anchors (eg, "POI," "defendant" for POIs and "victim," and "vic" for victim) including cases where:

- unspecified mental disorders were recorded simply as "the defendant has *mental health issues*," "victim is suffering from a *severe mental disorder*";
- psychotropic drugs were used by the POI or victim (eg, "the victim takes *Valium*," "accused takes a number of *antidepressants*") that might indicate a mental illness categorized in 4 groups (antianxiety, antidepressants, neuroleptics, antipsychotics);
- individual had traumatic brain injury, drug prescription abuse (unspecified in the text regarding the medication), substance abuse (unspecified in the text regarding the substance), and drug-induced disorders (unspecified in the text regarding the drug; see Multimedia Appendix 2 for the full reference list including our own 8 categories).

These rules were combined with dictionaries of terms for mental illness including common abbreviations and synonyms. The methodology was fully evaluated against the manual annotations of mental illness mentions for POIs and victims by 2 experts (in DV and neuropsychiatry, respectively) in a random sample of 100 police-recorded DV events, and returned an average 92% precision (ie, the percentage of correctly identified mental illness mentions against the total number of identified mental illness mentions, a denominator that includes both true-positive and false-positive mentions of mental illness by text mining) for the extracted mental illness mentions for both POIs (97.5% precision) and victims (87.1% precision); a detailed description has been published elsewhere [39].

The extracted mental illness mentions based on the description provided in the police records ranged from general descriptions (eg, "mood disorder," "behavioral problems") to very specific mentions (eg, "oppositional defiance disorder," "paranoid schizophrenia"). To impose a suitable structure for further analysis of the mental health data, we mapped the mental illness mentions to the World Health Organization's International Classification of Diseases (ICD-10) Mental and Behavioural Disorders categories using 4 levels (Multimedia Appendix 3) [41]. We utilized the expertise of the fourth author (PS) in the field of neuropsychiatry in cases where the mapping was ambiguous. The first level of mapping included 18 categories based on the ICD-10 with 8 additional customized categories; 4 categories where no specific disorder was mentioned but mental illness was implied by mention of a particular medication (see 22-25, Multimedia Appendix 2). This included specific drug classes such as antidepressants or specific brand names such as Zoloft. Four additional categories were added covering "drug prescription abuse," "substance abuse (unspecified)," "traumatic brain injury," and "unspecified drug induced disorders." Cases in which we recognized that either the victim or the POI had an unknown mental illness, or an unknown drug-induced mental disorder, were assigned into the categories of "unspecified mental disorder" and "unspecified drug induced disorder," respectively.

Cases in which mental illness mentions were more specific were mapped to lower-level ICD-10 categories. For example, the mention of “acute stress reaction” was classified at the third level according to the ICD-10 schema. Because the mention had a third-level mapping, this indicates that it can also be mapped upward to the second ICD-10 level (Reaction to severe stress and adjustment disorders) and first ICD-10 level, respectively (Anxiety, dissociative, stress-related, somatoform, and other nonpsychotic mental disorders).

In some narratives, a fourth level of detail based on the ICD-10 classification containing 27 categories was recorded by the police. However, for the purpose of reporting in this paper, we combined the fourth and third levels, for example, instead of reporting “other impulse disorders” (third level), we included “intermittent explosive disorder” (fourth level) in the third classification level (Multimedia Appendix 3 shows some mapped examples of extracted mental illness to ICD-10). For reporting purposes, we show only police-recorded DV events that involved POIs and victims with mental illness at the second level of ICD-10 because the first-level ICD-10 descriptors are too broad (eg, mood [affective] disorders).

Despite utilizing 2 sets of rules that correctly identified whether a mental illness mention was linked to a POI or a victim within a DV event, this text mining methodology was unable to associate the extracted mental illness “mention” with a specific POI or victim, if more than 2 individual POIs or victims were present. Thus, we focused only on those DV events that included a single POI and a single victim which included a total of 416,441 DV events out of 492,393. In this analysis results are presented for 416,441 police-recorded DV events. Seven age groups were defined (0-14 years, 15-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, and 65 years and over) to align with the national reporting practices of the Australian Institute of Health and Welfare’s Family, Domestic and Sexual Violence in Australia [42].

Results

Study Analysis

Overall, 15.51% of police-recorded DV events (64,587/416,441) had at least one mention of a mental illness for either the POI, victim, or both. In almost three-quarters (49,154/64,587, 76.11%) of these events, there was only a mental illness mention for the POI, 17.29% (11,167/64,587) for the victim, and 6.61% (4266/64,587) for both the victim and POI (Table 1). The total number of mental illness mentions classified at level 1 was 85,880 (100%); level 2, 51,858/85,880 (60.38%); and level 3, 26,704/85,880 (31.09%; Table 1). It should be noted that 1 police-recorded DV event can have more than 1 (different) mental illness mentions associated with a POI or victim. This can be seen in Table 1, where the total number of mental illness mentions is greater than the number of police-recorded DV events with a mental illness for a POI or a victim.

Table 1. Numbers of police-recorded DV events with mental illness mentions for POIs and victims (first column), and numbers of the mental illness for POIs and victims at various levels of the International Classification of Diseases-10 groupings.

Individual type	Police-recorded DV[a] events, n	Mental illness mentions (third level), n	Mental illness mentions (second level), n	Mental illness mentions (first level), n
POI[b] only	49,154	20,452	39,269	67,582
Victim only	11,167	6252	12,589	18,298
POI and victim	4266	N/A[c]	N/A	N/A
Total	64,587	26,704	51,858	85,880

[a] DV: domestic violence.

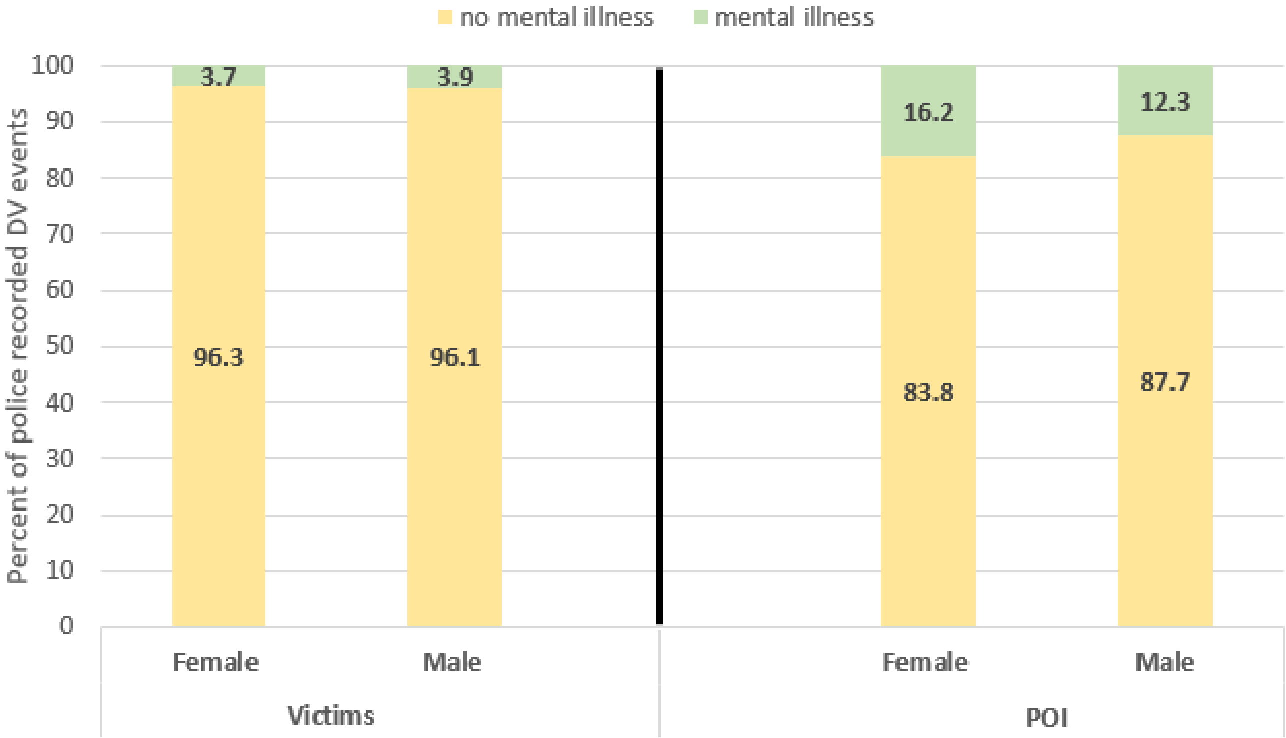
[b] POI: persons of interest.

[c] N/A: not applicable.

As police data are collected at the police-recorded DV event level, a single victim or POI may have multiple events over time. Of the 416,441 police-recorded events, the total number of unique POIs with a recorded mental illness was 18.53% (39,688/214,185) with 5.61% (13,709/244,219) for unique victims. This was lower than the current estimated prevalence from the Australian Bureau of Statistics for 2017-18, which revealed that 20.1% of the Australian population had a current mental or behavioral condition [43].

The percentage of police-recorded DV events in which victims had a mention of mental illness was lower than for POIs for both women and men (3.70% [11,523/311,210] and 3.86% [3718/96,228] vs 16.21% [12,048/74,323] and 12.30% [40,514/329,306]; Figure 1). Interestingly, 16.21% (12,048/74,323) of police-recorded DV events with female POIs had a recorded mental illness compared with 12.30% for men (40,514/329,306; Figure 1).

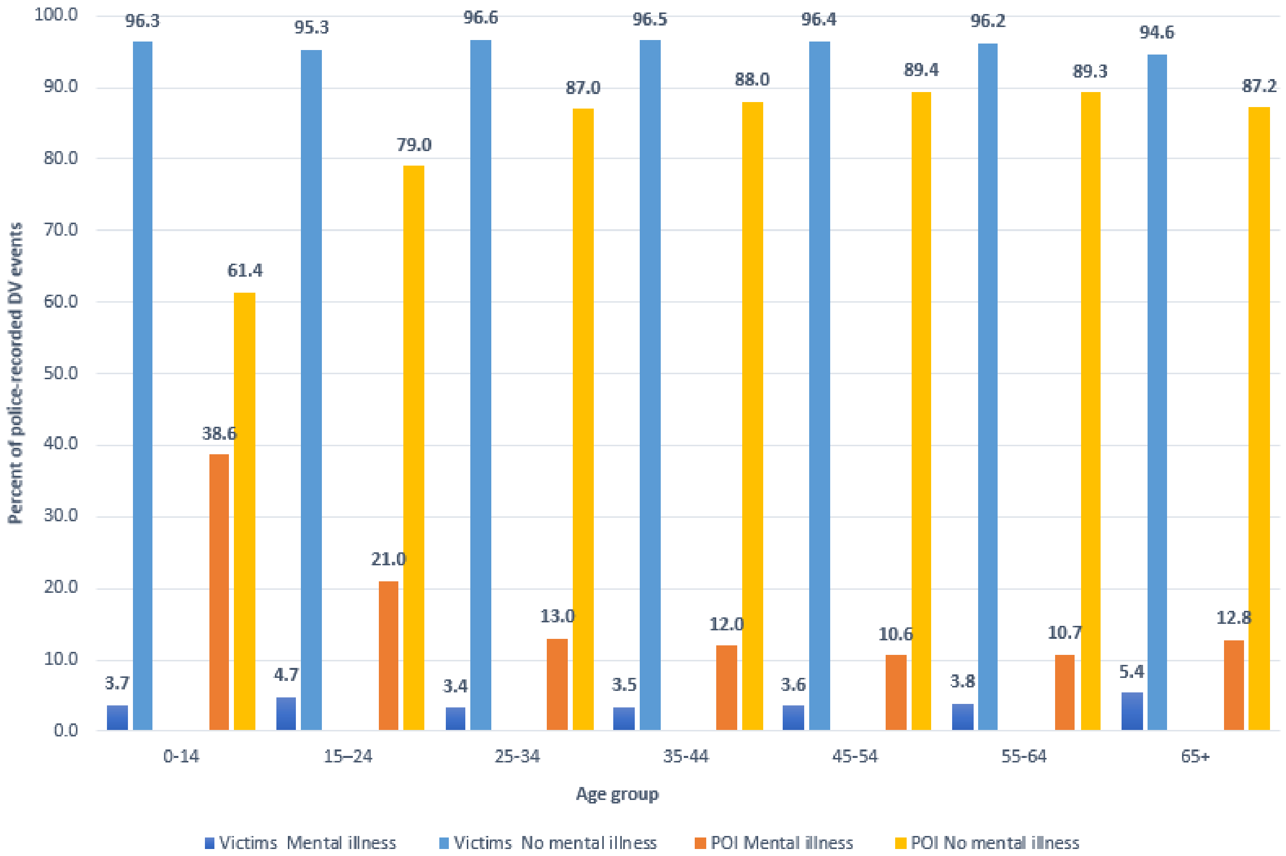
Figure 1. Proportion of police-recorded DV events with mental illness by sex of victims and POIs.



It is important to present “sampling errors” associated with the point estimates using confidence intervals. However, in our study 95% CIs were extremely narrow due to the large sample sizes. For example, the point estimate for the proportion of mental illness among female victims was 3.70% (11,523/311,210) with 95% CI of 3.6%-3.8%, indicating that 3.70% (11,523/311,210) was estimated with less than ±0.1% precision. Similarly, the proportion of mental illness among male victims was 3.86% (3718/96,228) with 95% CI of 3.7-4.0 (ie, 3.86% [3718/96,228] was estimated with less than ±0.1% precision). Extremely narrow intervals were also observed for the POIs—proportion of mental illness for females: 16.21% (12,048/74,323) with 95% CI of 15.9%-16.5%; proportion of mental illness for males: 12.30% (40,514/329,306) with 95% CI of 12.2-12.4.

When looking at the proportion of police-recorded DV events with mental illness by age group, the highest proportion of police-recorded DV events with a mental illness among victims was for the 65 years and over age group (1227/22,873, 5.36%; Figure 2). However, for POIs, the youngest age group showed the highest proportion of police-recorded DV events with a mental illness (623/1612, 38.64%, for POIs aged 0-14 years; Figure 2).

Figure 2. Percentage of police-recorded DV events with POIs and victims with or without a mental illness by age group.



Mental Illness Reported in ICD Levels 1 and 2

At the first ICD classification level (Table 2), almost one-third of the mentions of mental illness (22,172/67,582, 32.81%) for POIs and one-fifth (4208/18,298, 23.00%) for victims were classified as "unspecified mental illness," meaning the type of mental illness was not explicitly recorded in the narratives by the attending police officer(s). "Mood [affective] disorders" (eg, bipolar disorder, depression) had the highest number of mentions among POIs (12,753/67,582, 18.87%) and victims (4288/18,298, 23.43%), with "mental and behavioral disorders due to psychoactive substance use" (including alcohol abuse) ranking fourth and fifth for POIs (5642/67,582, 8.35%) and victims (1098/18,298, 6.00%), respectively. For POIs, 11.45% (7735/67,582) had mentions of "behavioral and emotional disorders with their onset usually occurring in childhood and adolescence" (eg, "attention deficit hyperactivity disorders," "conduct disorders"), whereas for victims this was 9.77% (1787/18,298). Although mentions of intellectual disabilities among POIs (n=1276) were higher than in the victims (n=813), as a percentage (of police-recorded DV events among POIs and among victims separately) this proportion was higher in victims (813/18,298, 4.44%) than POIs (1276/67,582, 1.89%). Mentions of traumatic brain injury (eg, "the victim has suffered a brain injury due to a car accident") were reported for 0.84% of POIs and 1.10% victims (568/67,582 and 201/18,298 mentions, respectively).

In the second-level categories (Table 3), "major depressive disorder, single episode" was the most commonly reported mental illness for both victims (2822/12,589, 22.42%) and POIs (7496/39,269, 19.01%), while alcohol abuse was the second highest mental illness among POIs (4867/39,269, 12.39%) and the fifth highest reported among victims (1033/12,589, 8.21%), further supporting the association between DV and alcohol use [44]. Additionally, there were 576 police-recorded DV events with victims (576/12,589, 4.58%) and 486 police-recorded DV events with POIs (486/39,269, 1.24%) with dementia. At the third level, "bipolar disorder, unspecified" was the most prevalent for both POIs (4462/20,452, 22.82%) and victims (1362/6252, 21.79%) with similar rates, followed by "unspecified behavioral and emotional disorders" (3930/20,452, 19.22%) for POIs and "anxiety disorder, unspecified" (1274/6252, 20.38%) for victims.

Table 2. Number of police-recorded DV events (%) with a mental illness mention grouped by the first ICD-10 level for POIs and victims from 416,441 DV events recorded by the New South Wales Police Force in Australia between 2005 and 2016.

Mental illness (ICD-10[a] code)	Police-recorded DV[b] event frequency (POI[c]), n (%) (N=67,582)	Prevalence, %[d]	Police-recorded DV event frequency (victim), n (%) (N=18,298)	Prevalence, %[d]
Unspecified mental disorder (F99)	22,172 (32.81)	5.32	4208 (23.00)	1.01
Mood (affective) disorders (F30-39)	12,753 (18.87)	3.0.6	4288 (23.43)	1.02
Behavioral and emotional disorders with onset usually occurring in childhood and adolescence (F90-99)	7735 (11.45)	1.85	1787 (9.77)	0.42
Mental and behavioral disorders due to psychoactive substance use (F10-19)	5642 (8.35)	1.35	1098 (6.00)	0.2.6
Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders (F20-29)	4751 (7.03)	1.14	893 (4.88)	0.21
Anxiety, dissociative, stress related, somatoform, and other nonpsychotic mental disorders (F40-49)	3034 (4.49)	0.72	1961 (10.72)	0.47
Intentional self-harm (X71-83)	2702 (4.00)	0.64	821 (4.49)	0.19
Substance abuse	2310 (3.42)	0.55	314 (1.72)	<0.1
Pervasive and specific developmental disorders (F80-89)	1492 (2.21)	0.35	417 (2.28)	0.10
Intellectual disability (F70-79)	1276 (1.89)	0.30	813 (4.44)	0.19
Disorders of adult personality and behavior (F60-69)	1096 (1.62)	0.26	369 (2.02)	<0.1
Injury of unspecified body region (T14)	687 (1.02)	0.16	221 (1.21)	<0.1
Traumatic brain injury	568 (0.84)	0.13	201 (1.10)	<0.1
Mental disorders due to known physiological conditions (F01-09)	498 (0.74)	0.11	580 (3.17)	0.13
Medication—antidepressants	326 (0.48)	<0.1	114 (0.62)	<0.1
Symptoms and signs involving cognition, perception, emotional state, and behavior (R40-46)	168 (0.25)	<0.1	74 (0.40)	<0.1
Medications—antipsychotics	108 (0.16)	<0.1	13 (0.07)	<0.1
Medications—antianxiety	77 (0.11)	<0.1	20 (0.11)	<0.1
Other degenerative diseases of the nervous system (G30-32)	54 (0.08)	<0.1	47 (0.26)	<0.1
Chromosomal abnormalities, not elsewhere classified (Q90-99)	48 (0.07)	<0.1	32 (0.17)	<0.1
Unspecified drug-induced disorders	43 (0.06)	<0.1	0 (0.00)	<0.1
Behavioral syndromes associated with physiological disturbances and physical factors (F50-59)	24 (0.04)	<0.1	18 (0.10)	<0.1
Systematic atrophies primarily affecting the central nervous system (G10-14)	11 (0.02)	<0.1	6 (<0.03)	<0.1
Diseases of the nervous system	3 (<0.01)	<0.1	3 (<0.01)	<0.1
Drug prescription abuse	3 (<0.01)	<0.1	0 (<0.01)	<0.1
Medication—neuroleptics	1 (<0.01)	<0.1	0 (<0.01)	<0.1

[a] ICD: International Classification of Diseases.

[b] DV: domestic violence.

[c] POI: persons of interest.

[d] Prevalence was calculated by dividing the number of police-recorded DV events with an ICD level 1 mention with the total number of police-recorded DV events.

Table 3. The 20 most commonly reported mental illnesses for both POIs and victims (at the second and third level of the ICD-10 categories) in 416,441 DV events recorded by the New South Wales Police Force in Australia between 2005 and 2016.

ICD-10[a] level 2			ICD-10 level 3		
Mental illness (ICD-10 code)	Police-recorded DV[b] events (POI[c]), n	Police-recorded DV events (victim), n	Mental illness (ICD-10 code)	Police-recorded DV events (POI), n	Police-recorded DV events (victim), n
Major depressive disorder, single episode (F32)	7496	2822	Bipolar disorder, unspecified (F31.9)	4462	1362
Alcohol abuse (F10)	4867	1033	Unspecified behavioral and emotional disorders with onset usually occurring in childhood and adolescence (F98)	3930	636
Bipolar disorder (F31)	4487	1361	Schizophrenia, unspecified (F20.9)	3762	711
Schizophrenia (F20)	4032	732	Anxiety disorder, unspecified (F41.9)	1910	1274
Other behavioral and emotional disorders with onset usually occurring in childhood and adolescence (F98)	3939	636	Autism (F84.0)	768	277
Attention deficit hyperactivity disorder (F90)	3045	1041	Suicide attempt (T14.91)	686	221
Other anxiety disorders (F41)	2001	1494	Cyclothymic disorder (F34.0)	633	71
Pervasive developmental disorder (F84)	1444	394	Post-traumatic stress disorder (F43.1)	619	327
Specific personality disorders (F60)	1074	329	Oppositional defiant disorder (F91.3)	614	92
Intellectual disability, unspecified (F79)	1023	669	Asperger syndrome (F84.5)	588	117
Conduct disorders (F91)	687	99	Paranoid personality disorder (F60.0)	538	137
Injury of unspecified body region (T14)	687	221	Personality disorder, unspecified (F60.9)	236	88
Persistent mood disorder (F34)	641	71	Obsessive compulsive disorder, unspecified (F42.9)	234	66
Reaction to severe stress, and adjustment disorders (F43)	638	335	Postpartum depression (F32.9)	227	233
Unspecified psychosis not due to a substance or known physiological condition (F29)	512	104	Borderline personality disorder (F0.3)	219	87
Dementia, unspecified (F03)	486	576	Paranoid schizophrenia[d] (F20.0)	207	23
Other psychoactive substance-related disorders[d] (F19)	291	19	Suicidal ideations (R45.85)	161	74
Obsessive-compulsive disorder (F42)	234	66	Dissociative identity disorder (F44.81)	114	38
Other stimulant related disorders[d] (F15)	213	16	Panic disorder (F41.0)	85	218
Cannabis abuse[d] (FF12)	174	12	Conduct disorder, unspecified[d] (F91.9)	72	7

ICD-10[a] level 2			ICD-10 level 3		
Mental illness (ICD-10 code)	Police-recorded DV[b] events (POI[c]), n	Police-recorded DV events (victim), n	Mental illness (ICD-10 code)	Police-recorded DV events (POI), n	Police-recorded DV events (victim), n
Intellectual disability, mild[e] (F70)	133	73	Alzheimer disease, unspecified[e] (G30.9)	46	46
Symptoms and signs involving emotional state[e] (R45)	168	72	Down syndrome, unspecified[e] (Q90.9)	48	32

[a] ICD: International Classification of Diseases.

[b] DV: domestic violence.

[c] POI: persons of interest.

[d] Top 20 common mental illnesses—POIs only.

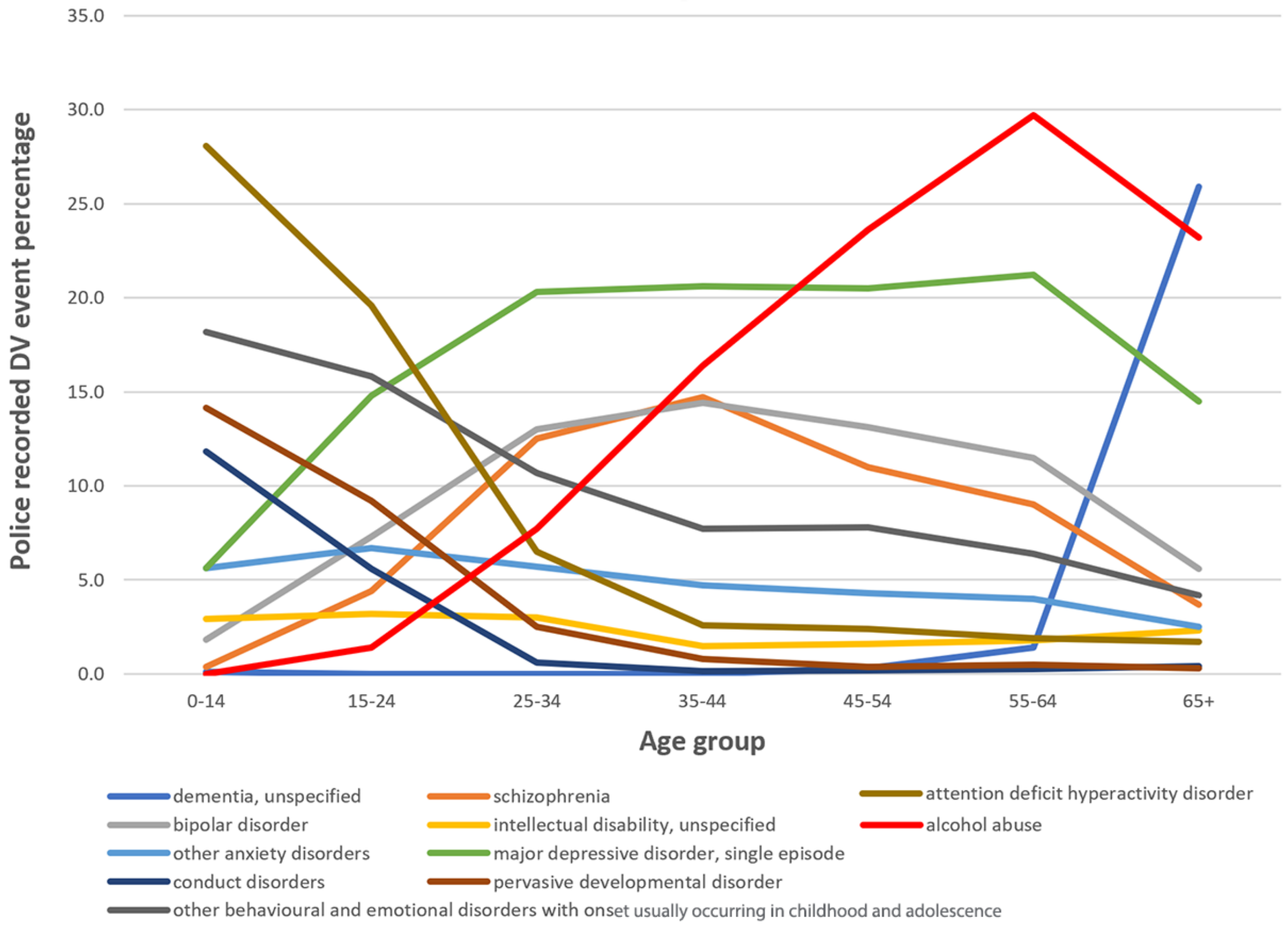
[e] Top 20 common mental illnesses—victims only.

Mental Illness by Age

Persons of Interest

The proportion of police-recorded DV events for POIs, "alcohol abuse" showed an increase from 15-24 years across all older age groups with the highest proportion in the 55-64 years group (695/2340, 29.70%; Figure 3). The most commonly reported mental illness for those who were 65 years and over was "dementia, unspecified" (320/1235, 25.91%).

Figure 3. Police-recorded DV event percentages for the top ten commonly reported mental illnesses across age groups for POIs.



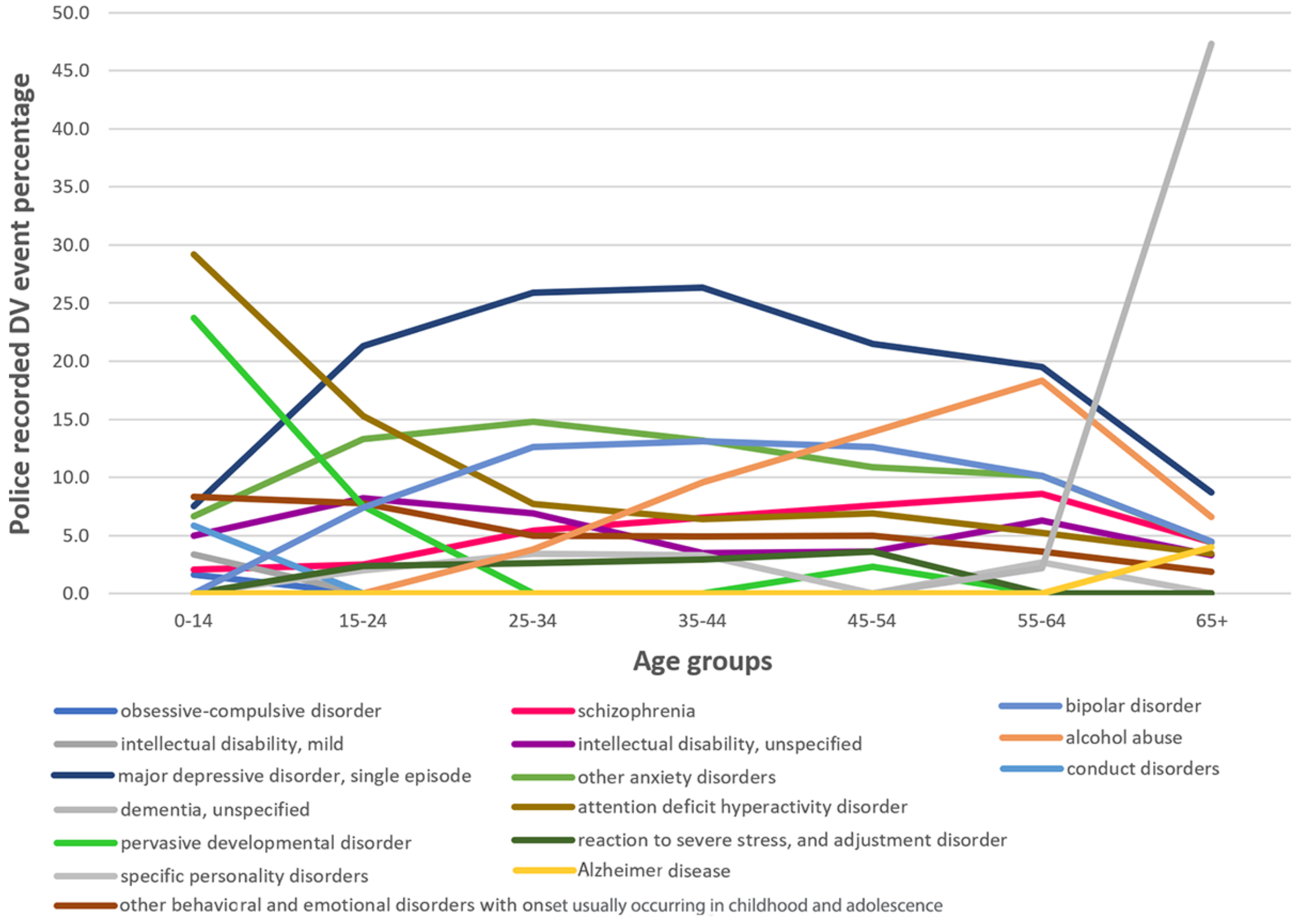
By contrast, "attention deficit hyperactivity disorder" was the most prevalent mental illness among the younger age groups making up 28.1% (230/818) and 19.60% (1657/8454) of police-recorded DV events for the 0-14 years and 15-24 years age groups, respectively. Mental illness associated with younger populations such as "pervasive development disorder" and "conduct disorders" made up a relatively high proportion of police-recorded DV events for their age groups (116/816 [14.2%] and 97/822 [11.8%], respectively). Mental illness in the younger age groups showed a decline with increasing age (eg, “pervasive developmental disorders,” “conduct disorders,” “attention deficit hyperactivity disorder,” “other behavioral and emotional disorders with onset during childhood,” “other anxiety disorders,” and “intellectual disability, unspecified”).

“Major depressive disorder, single episode” was the most common mental illness in police-recorded DV events involving those aged 25-34 years and 35-44 years (1927/9492 [20.30%] and 1880/9126 [20.60%], respectively). “Bipolar disorder” and “schizophrenia” showed a similar trend, increasing as a proportion of their age group, until the 35-44 years age group, and then gradually decreasing for the older age groups.

Victims

Excluding the youngest and oldest age groups (ie, 0-14 and 65 years and over), “major depressive disorder, single episode” made up the greatest proportion of recorded mental illness across all other age groups ranging from 19.51% (224/1148; 55-64 years) to 26.30% (748/2844; 35-44 years). “Attention deficit hyperactivity disorder” made up 29.1% (70/240) of recorded mental illnesses for those aged 0-14 years and for victims aged 65 years and over, “dementia, unspecified” was recorded for almost half (515/1088, 47.33%) of police-recorded DV events with a reported mental illness (Figure 4).

Figure 4. Police-recorded DV event percentages for the top ten commonly reported mental illnesses across age groups for victims.



Conditions in the younger groups (0-14 and 15-24) such as “pervasive developmental disorder” and “conduct disorders” gradually decreased across older age groups. As a proportion of recorded mental illnesses, “major depressive disorder, single episode” and “bipolar disorder” increased from the younger age groups to their highest in the 35-44 year group (748/2844 [26.30%] and 374/2854 [13.10%], respectively), and then steadily decreased in the older age groups (95/1091 [8.71%] for “major depressive disorder, single episode” and 50/1111 [4.50%] for “bipolar disorder” in the 65 years and over old group). Additionally, “alcohol abuse” showed a steady increase across the age groups reaching a peak in the 55-64 years age group (210/1147, 18.31%). A similar trend was observed for “schizophrenia,” increasing to 8.71% (99/1137) of police-recorded DV events with a mental illness mention for the 55-64-year group. For the exact percentages for Figures 3 and 4, see Multimedia Appendices 4 and 5.

Discussion

Principal Findings

By text mining a large, population-based data set of DV events recorded by the NSWPF, our findings indicate that a large number of those events (64,587/416,441, 15.51%) involve individuals (victims and POIs) who may have a mental illness. The findings are important in raising awareness about the significance of mental illness in the context of DV and have implications for the training of front-line police officers in managing those with mental illness attending DV events. While there is a growing literature concerned with the association between mental illness and DV, no other published study, as far as we are aware, has reported on the real-time capture of mental illness data by front-line police officers attending and recording DV events.

These findings complement previously published research indicating that mental illness can increase the risk of being in an abusive relationship, as either a POI or victim (or both) [3,14,20,45]. Further research is required to explore the unique context in which this arm (the police) of the justice system interacts with individuals with mental illness and how this can be optimized to improve outcomes in these situations. Police officer detection of possibly relevant mental illnesses that are virtually contemporaneous when attending to a DV event can lead to the identification of potential new strategies and interventions to tackle this issue. An example could be the development of a mobile app utilized by the police who by recording a mental health diagnosis can receive guidelines to de-escalate the DV situation or improve their decision making.

Our results showed that the prevalence of mental illness for unique victims (13,709/244,219, 5.61%) and POIs (39,688/214,185, 18.53%), respectively, was lower than that of the estimated national prevalence of mental illness reported in the 2017-18 National Health Survey (20.1%, 4.8 million) [43]. There are several possible reasons as to why our estimated prevalence was lower. Our data were from 2005 to 2016, and rates of mental illness may have increased over time. The National Health Survey sets out to systematically capture prevalent mental illness. By contrast, underreporting of mental illness to the police in these events is to be anticipated given that the detection and recording of mental health status are not the focus of the police's work since their role is not to diagnose or inquire about mental illness but to prioritize victim safety and diffuse the situation. Most likely, mental illness is reported in a very ad hoc manner and only if the POI, victim, or other person divulges this information.

Our estimates of mental illness were derived from unique persons from single victim to single POI police-recorded DV event data, while police-recorded DV events with multiple POIs or victims were excluded (and therefore some mental illness-related information was lost). The National Health Survey's self-reported mental and behavioral conditions information may encompass a broader definition of mental health and well-being than what was captured in the police narratives [43]. It is noteworthy that the difference in the overall rates of mental illness mentions found in this study—lower in victims than in POIs—does not seem to be reflected in the existing literature. It is possible that a bias exists in the context of police questioning on the mental health status of the POIs such that relatively more mental illnesses are reported for this group. Such a bias, if confirmed, would have potentially important implications, particularly if the detection of mental illness by police were to influence the provision of immediate support. Finally, false negatives (correct mentions of mental illness for POIs and victims ignored by the rules) generated by the application of text mining could potentially have contributed to this low prevalence, prompting to cast a wider net of rules that could capture more generic mentions and avoid the reliance on semantic anchors.

We found differences in the top 10 most commonly reported mental illnesses across age groups for POIs and victims. While the top 10 most reported mental illnesses among POIs remained consistent across age groups, this did not occur for victims. For example, "obsessive compulsive disorder" and "intellectual disability, mild" were only reported once among younger age groups. "Alzheimer disease, unspecified" was one of the most commonly reported illnesses for the 65 years and over age group for victims.

Given the low representation of personality disorder in our data, while this being a common diagnosis for DV perpetrators in published studies, it is plausible that personality disorder might account for a significant proportion of unspecified mental illness, and more so in perpetrators than victims [46].

Studies have consistently found heavy alcohol use to be associated with DV in both men and women [47,48]. In NSW, the Australian Institute of Criminology has shown evidence for alcohol misuse as an important risk factor for DV [49]. Our results indicated that the number of police-recorded DV events among POIs involving "alcohol abuse" increased with increasing age, with the highest number of police-recorded DV events shown to be among POIs of 55-64 years old. These findings, at population-level sample, support the link between alcohol abuse and DV.

Conditions that (usually) occur in childhood such as "attention deficit hyperactivity disorder" or "conduct disorders" were understandably most prevalent in the younger age groups. Individuals with these conditions are potentially vulnerable to domestic abuse [50] as well as at risk of committing violence toward parents, peers, or carers, likely reflecting the fact that impaired behavioral self-regulation implied by these diagnoses serves as a risk factor for both aggressive acts and reaction from others in response to what might be perceived as provocation [51].

We observed an increase in police-recorded DV events with dementia among the older age groups for both POIs and victims. The plethora of evidence suggests that older individuals with dementia are at a high risk of abuse, especially in a carer setting [52,53]. Our findings add to this evidence base, showing that among victims aged 65 years and over, dementia was implicated in 47.33% (515/1088) of police-recorded DV events.

DV has been directly linked to severe mental illnesses including mood disorders [3]. Depression, in particular, has been associated with both victimization and perpetration of DV, with the extent of abuse corresponding with the severity of depression [54,55], something that has been reflected in our findings in the early and mid-adulthood groups [54].

Schizophrenia was proportionally higher among younger POIs (25-44 years old) and older victims (55-64 years old; 12.50%-14.70% [1188/9504-1346/9156] and 8.60% [99/1151], respectively). Our findings concur with previous studies suggesting that those with schizophrenia have a higher risk of perpetrating violence [56] and DV toward family members [57]. Previous studies also suggest that individuals with schizophrenia can be vulnerable in a domestic setting and open to experience more types of victimization [58,59].

In contrast to the police-recorded DV events with mental illness mentions identified using text mining, the structured data in the WebCOPS system contained a field entitled “mental illness related.” The total number of police-recorded DV events flagged as “mental illness related” was 1.03% (n=4295) of the total number of police-recorded DV events (N=416,441). This is in contrast to the number of police-recorded DV events that had extracted mental illness mentions from the narratives for the same events amounting to a total of 64,587 (15.51%). This discrepancy is likely explained by the police making a judgment call that mental illness was not considered as a factor for the cause of a DV event. However, through the application of text mining, we identified almost 16 times more police-recorded DV events with mental illness implicated than the police had classified in the fixed field as “mental illness related.” Further investigation to determine how these judgment calls are made by the police and the benefit of making this determination is required.

Automatically extracting mental illness mentions can add to existing data regarding POIs and victims involved in DV events and potentially in future events altering the police’s response toward a person with a known mental illness. One practical application of extracting this type of information can be its use in models along with other identified features (eg, abuse types, victim injuries) that could predict future offences within the area of DV and utilizing machine learning approaches, which could enable improved allocation of police sources for DV management. The successful implementation of text mining in police-recorded DV events may encourage greater use of unstructured data within law enforcement agencies that can be processed by such automated methodologies to extract important information regarding DV and other types of recorded offences (eg, sexual abuse, child neglect) with the police. The study has demonstrated that the trove of information contained in these events can be used to raise awareness among police officers regarding mental illness and, alongside better training, can improve the management of DV cases involving individuals with mental illness. With improved identification and awareness, it provides options for the police to divert individuals to hospital or community mental health services as appropriate. We believe this long-term preventative jurisprudence approach may provide opportunities to respond appropriately to mental illness in police-recorded DV events.

Limitations

We cannot be certain that any individual extracted mention of a mental illness from police-recorded DV events is accurate. Police officers do undergo mental health training in 1- to 4-day courses (NSWPF, personal communication) so they can be aware if a POI or a victim may have a mental illness in addition to being informed by the victims and the POIs themselves, or potential witnesses of the event, and based on the evidence in the scene (eg, presence of medication prescriptions, drug and alcohol use). However, no literature exists on the validity of self-reported psychiatric status when shared with health professionals, or in other contexts where the data may be sought for administrative reasons, let alone with the police officers attending a home following a highly charged DV event. Studies reflecting self-diagnosed mental health conditions have demonstrated low validity with a substantial underreporting of mental health issues, which could be a reflected effort to avoid stigma [60,61].

Future work should examine the veracity of the police mentions of mental illness by using formal diagnostic information available from administrative data collections in hospital admissions, GP presentations, and community mental health services. It would be particularly interesting to see if mentions of diagnoses with high implied precision (ie, second- and particularly third-level diagnostic categories) are more likely to be validated by existing records than the more generic diagnoses. Such a study would allow a determination as to whether there is a bias in respect of greater police mentions of mental illness for POIs instead for the victims. In addition, further exploration will be conducted to investigate the observed differences between most commonly reported mental illnesses across the age groups of POIs and victims as well as to investigate whether the extracted information can be used as input toward predictive models for DV.

Conclusions

This novel study involving the automated extraction of mental illness mentions through text mining from a large-scale data set of 416,441 police-recorded DV events provides potentially important information for mental health professionals and criminal justice policy makers to help address mental illness in police-recorded DV events. A trove of DV data are captured by the police as unstructured text that text mining can unearth. The information extracted from a large-scale set of police-recorded DV events suggests there may be more in-depth information related to trends in mental illness for victims and POIs. While this information can be seen as police insights in recorded DV events, it can provide the basis for examining the concordance of the extracted mental illness mentions with official diagnosis from health records and research that aims to assess the characteristics and features of victims and POIs involved in police-recorded DV events. This work will also explore whether extracted information can be used to design predictive models for those at risk of further victimization, to inform prevention strategies that could be implemented at the early stages of police involvement in a DV event.

Acknowledgments

The authors thank the following individuals from the NSWPF for assistance in accessing the data and advice regarding police procedures: Dr Chris Devery, Dr Christie Wallace, John Blanchette, Erin Sharland, and Nicole Grant. This research was supported by an Australian Institute of Criminology/Criminology Research Grant (34/15-16).

Conflicts of Interest

None declared.

Multimedia Appendix 1
A hypothetical example of a domestic violence event narrative as recorded by the NSWPF.
[DOCX File , 83 KB - jmir_v22i12e23725_app1.docx]

Multimedia Appendix 2
Mental illnesses listed in the ICD-10 including the eight new categories targeted for extraction in domestic violence events with examples as they appeared in the police event narratives.
[DOCX File , 16 KB - jmir_v22i12e23725_app2.docx]

Multimedia Appendix 3
The ICD-10 Mental and Behavioural Disorders schema used to map the extracted mental illness mentions containing three levels (first, second and third).
[DOCX File , 39 KB - jmir_v22i12e23725_app3.docx]

Multimedia Appendix 4
Percentage of domestic violence events involving POIs with the top ten most commonly mentioned mental illnesses at ICD-10 level 2 across age groups.
[DOCX File , 15 KB - jmir_v22i12e23725_app4.docx]

Multimedia Appendix 5
Percentage of domestic violence events involving victims with the top ten most commonly mentioned mental illnesses at ICD-10 level 2 across age groups.
[DOCX File , 16 KB - jmir_v22i12e23725_app5.docx]

References

1. Home Office. Crime in England and Wales: Findings From the British Crime Survey and Police Recorded Crime. London, UK: Home Office; 2009. URL: http://news.bbc.co.uk/2/shared/bsp/hi/pdfs/17_07_08_crime_statistics_200708.pdf [accessed 2020-12-14]
2. Briodi A. Domestic Violence is a crime [electronic resource]: police and community booklet. Sydney, NSW: Sydney City Council and NSWPF; 2010. URL: http://nla.gov.au/nla.arc-121383 [accessed 2019-01-10]
3. Howard LM, Trevillion K, Khalifeh H, Woodall A, Agnew-Davies R, Feder G. Domestic violence and severe psychiatric disorders: prevalence and interventions. Psychol. Med 2009 Nov 06;40(6):881-893. [doi: 10.1017/s0033291709991589]
4. Trevillion K, Oram S, Feder G, Howard LM. Experiences of domestic violence and mental disorders: a systematic review and meta-analysis. PLoS One 2012 Dec 26;7(12):e51740 [FREE Full text] [doi: 10.1371/journal.pone.0051740] [Medline: 23300562]
5. Australian Institute of Health and Welfare. Family, Domestic and Sexual Violence in Australia 2018. Canberra, ACT: Australian Institute of Health and Welfare; 2018. URL: https://www.aihw.gov.au/reports/domestic-violence/family-domestic-sexual-violence-in-australia-2018/contents/summary [accessed 2019-01-10]
6. Bryant B, Cussen T. Homicide in Australia: 2010-11 to 2011-12: National Homicide Monitoring Program Report. Canberra, ACT: Australian Institute of Criminology; 2017.
7. Campo M. Children's exposure to domestic and family violence: Key issues and responses (CFCA Paper No. 36). Melbourne, VIC: Australian Institute of Family Studies; 2015:1-24.
8. Khalifeh H, Moran P, Borschmann R, Dean K, Hart C, Hogg J, et al. Domestic and sexual violence against patients with severe mental illness. Psychol. Med 2014 Sep 04;45(4):875-886. [doi: 10.1017/s0033291714001962]
9. The cost of violence against women and their children in Australia 2016. 2016 Aug. URL: https://www.dss.gov.au/sites/default/files/documents/08_2016/the_cost_of_violence_against_women_and_their_children_in_australia_-_summary_report_may_2016.pdf [accessed 2019-10-01]
10. Holmes MR, Richter FGC, Votruba ME, Berg KA, Bender AE. Economic Burden of Child Exposure to Intimate Partner Violence in the United States. J Fam Viol 2018 Feb 19;33(4):239-249. [doi: 10.1007/s10896-018-9954-7]
11. Oliver R, Alexander B, Roe S, Wlasny M. The economic and social costs of domestic abuse. London: Home Office; 2019. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/918897/horr107.pdf [accessed 2020-01-10]

12. Khalifeh H, Oram S, Trevillion K, Johnson S, Howard LM. Recent intimate partner violence among people with chronic mental illness: findings from a national cross-sectional survey. Br J Psychiatry 2015 Sep 02;207(3):207-212 [FREE Full text] [doi: 10.1192/bjp.bp.114.144899] [Medline: 26045349]
13. Choe JY, Teplin LA, Abram KM. Perpetration of violence, violent victimization, and severe mental illness: balancing public health concerns. Psychiatr Serv 2008 Feb;59(2):153-164 [FREE Full text] [doi: 10.1176/ps.2008.59.2.153] [Medline: 18245157]
14. Van Dorn R, Volavka J, Johnson N. Mental disorder and violence: is there a relationship beyond substance use? Soc Psychiatry Psychiatr Epidemiol 2012 Mar 26;47(3):487-503. [doi: 10.1007/s00127-011-0356-x] [Medline: 21359532]
15. Junginger J, McGuire L. Psychotic motivation and the paradox of current research on serious mental illness and rates of violence. Schizophr Bull 2004 Jan 01;30(1):21-30. [doi: 10.1093/oxfordjournals.schbul.a007064] [Medline: 15176759]
16. Oram S, Flynn SM, Shaw J, Appleby L, Howard LM. Mental illness and domestic homicide: a population-based descriptive study. Psychiatr Serv 2013 Oct;64(10):1006-1011. [doi: 10.1176/appi.ps.201200484] [Medline: 23820784]
17. Fazel S, Lichtenstein P, Grann M, Goodwin GM, Långström N. Bipolar disorder and violent crime: new evidence from population-based longitudinal studies and systematic review. Arch Gen Psychiatry 2010 Sep 01;67(9):931-938. [doi: 10.1001/archgenpsychiatry.2010.97] [Medline: 20819987]
18. Labrum T, Solomon P, Marcus S. Victimization and Perpetration of Violence Involving Persons With Mood and Other Psychiatric Disorders and Their Relatives. Psychiatr Serv 2020 May 01;71(5):498-501. [doi: 10.1176/appi.ps.201900384] [Medline: 32041513]
19. Fazel S, Gulati G, Linsell L, Geddes JR, Grann M. Schizophrenia and violence: systematic review and meta-analysis. PLoS Med 2009 Aug 11;6(8):e1000120 [FREE Full text] [doi: 10.1371/journal.pmed.1000120] [Medline: 19668362]
20. Khalifeh H, Dean K. Gender and violence against people with severe mental illness. Int Rev Psychiatry 2010 Nov 03;22(5):535-546. [doi: 10.3109/09540261.2010.506185] [Medline: 21047165]
21. Sariaslan A, Arseneault L, Larsson H, Lichtenstein P, Fazel S. Risk of Subjection to Violence and Perpetration of Violence in Persons With Psychiatric Disorders in Sweden. JAMA Psychiatry 2020 Apr 01;77(4):359-367 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.4275] [Medline: 31940015]
22. Bhavsar V, Dean K, Hatch SL, MacCabe JH, Hotopf M. Psychiatric symptoms and risk of victimisation: a population-based study from Southeast London. Epidemiol Psychiatr Sci 2018 Sep 10;28(2):168-178. [doi: 10.1017/s2045796018000537]
23. Van Deinse TB, Macy RJ, Cuddeback GS, Allman AJ. Intimate partner violence and sexual assault among women with serious mental illness: A review of prevalence and risk factors. Journal of Social Work 2018 Apr 11;19(6):789-828. [doi: 10.1177/1468017318766425]
24. Neria Y, Bromet EJ, Carlson GA, Naz B. Assaultive trauma and illness course in psychotic bipolar disorder: findings from the Suffolk county mental health project. Acta Psychiatr Scand 2005 May;111(5):380-383. [doi: 10.1111/j.1600-0447.2005.00530.x] [Medline: 15819732]
25. Macdonald W, Fitzgerald J. Understanding fraud: The nature of fraud offences recorded by NSW Police. BOCSAR NSW Crime and Justice Bulletins. 2014. URL: https://www.bocsar.nsw.gov.au/Pages/bocsar_publication/Pub_Summary/CJB/cjb180-Understanding-fraud-The-nature-of-fraud-offences-recorded-by-NSW-Police.aspx [accessed 2020-12-13]
26. Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic literature review. Int J Methods Psychiatr Res 2016 Jun 17;25(2):86-100 [FREE Full text] [doi: 10.1002/mpr.1481] [Medline: 26184780]
27. Spasić I, Livsey J, Keane JA, Nenadić G. Text mining of cancer-related information: review of current status and future directions. Int J Med Inform 2014 Sep;83(9):605-623 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.009] [Medline: 25008281]
28. Sohn S, Kocher J, Chute C, Savova G. Drug side effect extraction from clinical narratives of psychiatry and psychology patients. J Am Med Inform Assoc 2011 Dec;18 Suppl 1(Supplement_1):i144-i149 [FREE Full text] [doi: 10.1136/amiajnl-2011-000351] [Medline: 21946242]
29. Perlis RH, Iosifescu DV, Castro VM, Murphy SN, Gainer VS, Minnier J, et al. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. Psychol. Med 2011 Jun 20;42(1):41-50. [doi: 10.1017/s0033291711000997]
30. Eriksson R, Jensen PB, Frankild S, Jensen LJ, Brunak S. Dictionary construction and identification of possible adverse drug events in Danish clinical narrative text. J Am Med Inform Assoc 2013 Sep 01;20(5):947-953 [FREE Full text] [doi: 10.1136/amiajnl-2013-001708] [Medline: 23703825]
31. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register. PLoS One 2015 Aug 14;10(8):e0134208 [FREE Full text] [doi: 10.1371/journal.pone.0134208] [Medline: 26273830]
32. Karystianis G, Nevado AJ, Kim C, Dehghan A, Keane JA, Nenadic G. Automatic mining of symptom severity from psychiatric evaluation notes. Int J Methods Psychiatr Res 2018 Mar 22;27(1):e1602 [FREE Full text] [doi: 10.1002/mpr.1602] [Medline: 29271009]
33. Jackson RG, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, et al. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction

(CRIS-CODE) project. BMJ Open 2017 Jan 17;7(1):e012012 [FREE Full text] [doi: 10.1136/bmjopen-2016-012012] [Medline: 28096249]

34. Wu C, Kuo C, Su C, Wang S, Dai H. Using text mining to extract depressive symptoms and to validate the diagnosis of major depressive disorder from electronic health records. J Affect Disord 2020 Jan 01;260:617-623. [doi: 10.1016/j.jad.2019.09.044] [Medline: 31541973]
35. Chau M, Xu J, Chen H. Extracting meaningful entities from police narrative reports. In: Proceedings of the 2002 Annual National Conference on Digital Government Research. 2002 Presented at: Annual National Conference on Digital Government Research; May 19-22, 2002; Los Angeles, CA.
36. Ananyan S. Crime pattern analysis through text mining. In: AMCIS 2004 Proceedings. 2004 Presented at: Americas Conference on Information Systems; August 6-8, 2004; New York, NY p. 236.
37. Poelmans J, Elzinga P, Viaene S, Dedene G. Formally analysing the concepts of domestic violence. Expert Systems with Applications 2011 Apr;38(4):3116-3130. [doi: 10.1016/j.eswa.2010.08.103]
38. Haleem M, Han L, Harding P, Ellison M. An Automated Text Mining Approach for Classifying Mental-Ill Health Incidents from Police Incident Logs for Data-Driven Intelligence. 2019 Presented at: IEEE International Conference on Systems, Man and Cybernetics (SMC); October 7, 2019; Bari, Italy p. 2279-2284. [doi: 10.1109/smc.2019.8914240]
39. Karystianis G, Adily A, Schofield P, Knight L, Galdon C, Greenberg D, et al. Automatic extraction of mental health disorders from domestic violence police narratives: text mining study. J Med Internet Res 2018 Sep 13;20(9):e11548 [FREE Full text] [doi: 10.2196/11548] [Medline: 30213778]
40. Cunningham H, Tablan V, Roberts A, Bontcheva K. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. PLoS Comput Biol 2013 Feb 7;9(2):e1002854 [FREE Full text] [doi: 10.1371/journal.pcbi.1002854] [Medline: 23408875]
41. World Health Organization. The International Classification of Disease 10 of Mental and Behavioural Disorders. 2017. URL: https://www.who.int/substance_abuse/terminology/icd_10/en/ [accessed 2020-01-10]
42. Australian Institute of Health and Welfare. Family, domestic and sexual violence in Australia: continuing the national story. 2019. URL: https://www.aihw.gov.au/reports/domestic-violence/family-domestic-sexual-violence-australia-2019/contents/table-of-contents [accessed 2020-02-15]
43. Australian Bureau of Statistics. National Health Survey: First Results, 2017-18. 2019. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey-first-results/latest-release [accessed 2020-02-15]
44. Zilberman ML, Blume SB. [Domestic violence, alcohol and substance abuse]. Braz J Psychiatry 2005 Oct;27 Suppl 2(suppl 2):S51-S55 [FREE Full text] [doi: 10.1590/s1516-44462005000600004] [Medline: 16302054]
45. Maniglio R. Severe mental illness and criminal victimization: a systematic review. Acta Psychiatr Scand 2009 Mar;119(3):180-191. [doi: 10.1111/j.1600-0447.2008.01300.x] [Medline: 19016668]
46. Spencer C, Mallory AB, Cafferky BM, Kimmes JG, Beck AR, Stith SM. Mental health factors and intimate partner violence perpetration and victimization: A meta-analysis. Psychology of Violence 2019 Jan;9(1):1-17. [doi: 10.1037/vio0000156]
47. Livingston M. A longitudinal analysis of alcohol outlet density and domestic violence. Addiction 2011 May;106(5):919-925. [doi: 10.1111/j.1360-0443.2010.03333.x] [Medline: 21205052]
48. Sipilä M, Hakulinen T, Helminen M, Seppänen J, Paavilainen E, Koponen P. Alcohol abuse, psychological distress, and suicidal thoughts are associated with intimate partner violence among parents' with children. Mental Health & Prevention 2018 Dec;12:76-81. [doi: 10.1016/j.mhp.2018.09.006]
49. Morgan A, McAtamney A. Key issues in alcohol-related violence. Sydney, NSW: Australian Institute of Criminology; 2009.
50. Ebeling H, Nurkkala H. Children and adolescents with developmental disorders and violence. Int J Circumpolar Health 2002;61 Suppl 2:51-60. [doi: 10.3402/ijch.v61i0.17502] [Medline: 12585820]
51. Ilomäki E, Viilo K, Hakko H, Marttunen M, Mäkikyrö T, Räsänen P, STUDY-70 Workgroup. Familial risks, conduct disorder and violence: A Finnish study of 278 adolescent boys and girls. Eur Child Adolesc Psychiatry 2006 Feb;15(1):46-51. [doi: 10.1007/s00787-006-0507-x] [Medline: 16514509]
52. McCausland B, Knight L, Page L, Trevillion K. A systematic review of the prevalence and odds of domestic abuse victimization among people with dementia. Int Rev Psychiatry 2016 Oct 26;28(5):475-484. [doi: 10.1080/09540261.2016.1215296] [Medline: 27564566]
53. Frazão SL, Silva MS, Norton P, Magalhães T. Domestic violence against elderly with disability. J Forensic Leg Med 2014 Nov;28:19-24. [doi: 10.1016/j.jflm.2014.09.003] [Medline: 25440142]
54. Shorey RC, Febres J, Brasfield H, Stuart GL. The Prevalence of Mental Health Problems in Men Arrested for Domestic Violence. J Fam Violence 2012 Nov 01;27(8):741-748 [FREE Full text] [doi: 10.1007/s10896-012-9463-z] [Medline: 23284227]
55. Huss MT, Langhinrichsen-Rohling J. Assessing the generalization of psychopathy in a clinical sample of domestic violence perpetrators. Law Hum Behav 2006 Oct;30(5):571-586. [doi: 10.1007/s10979-006-9052-x] [Medline: 17031606]
56. Fazel S, Grann M, Carlström E, Lichtenstein P, Långström N. Risk factors for violent crime in Schizophrenia: a national cohort study of 13,806 patients. J Clin Psychiatry 2009 Mar 10;70(3):362-369. [doi: 10.4088/jcp.08m04274] [Medline: 19284931]

57. Kageyama M, Yokoyama K, Nagata S, Kita S, Nakamura Y, Kobayashi S, et al. Rate of Family Violence Among Patients With Schizophrenia in Japan. Asia Pac J Public Health 2015 Sep 16;27(6):652-660. [doi: 10.1177/1010539515595069] [Medline: 26182940]
58. Marley JA, Buila S. Crimes against people with mental illness: types, perpetrators, and influencing factors. Soc Work 2001 Apr 01;46(2):115-124. [doi: 10.1093/sw/46.2.115] [Medline: 11329641]
59. Honkonen T, Henriksson M, Koivisto A, Stengård E, Salokangas R. Violent victimization in schizophrenia. Soc Psychiatry Psychiatr Epidemiol 2004 Aug;39(8):606-612. [doi: 10.1007/s00127-004-0805-x] [Medline: 15300370]
60. Neumann AM, Giambrone AK, Yanes PK, Jaanimägi U, Wisniewski AM, Valtin LL, et al. Validity of self-reported psychiatric diagnoses among patients seeking treatment for a substance use disorder. Mental Health and Substance Use 2013 May;6(2):111-123. [doi: 10.1080/17523281.2012.693129]
61. Schofield P, Butler T, Hollis S, D'Este C. Are prisoners reliable survey respondents? A validation of self-reported traumatic brain injury (TBI) against hospital medical records. Brain Inj 2011 Nov 30;25(1):74-82. [doi: 10.3109/02699052.2010.531690] [Medline: 21117913]

Abbreviations

DV: domestic violence
GATE: General Architecture for Text Engineering
ICD: International Classification of Diseases
NSWPF: New South Wales Police Force
POI: persons of interest

Edited by G Eysenbach, R Kukafka; submitted 24.08.20; peer-reviewed by P Matthews, C Emezue; comments to author 16.09.20; revised version received 17.09.20; accepted 23.11.20; published 24.12.20.

Please cite as:
Karystianis G, Simpson A, Adily A, Schofield P, Greenberg D, Wand H, Nenadic G, Butler T
Prevalence of Mental Illnesses in Domestic Violence Police Records: Text Mining Study
J Med Internet Res 2020;22(12):e23725
URL: http://www.jmir.org/2020/12/e23725/
doi:10.2196/23725
PMID:33361056

Original Paper

Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions

Kimberly D Acquaviva[1], PhD, MSW; Josh Mugele[2], MD; Natasha Abadilla[3], BA; Tyler Adamson[4], MPH; Samantha L Bernstein[5], MSN, RNC-OB, IBCLC; Rakhee K Bhayani[6], MD; Annina Elisabeth Büchi[7], MSc, MD; Darcy Burbage[8], DNP, RN, AOCN, CBCN; Christopher L Carroll[9], MD; Samantha P Davis[10], MS, RRT, RRT-NPS, CHSE; Natasha Dhawan[11], MD; Alice Eaton[12], BSc; Kim English[13], RN, BScN, MN; Jennifer T Grier[14], PhD; Mary K Gurney[15], RPh, PhD, BCPA; Emily S Hahn[16], MD; Heather Haq[17], MHS, MD; Brendan Huang[18], BS; Shikha Jain[19], MD, FACP; Jin Jun[20], RN, PhD; Wesley T Kerr[21], MD, PhD; Timothy Keyes[3], BA; Amelia R Kirby[1], BA; Marion Leary[22], RN, MSN, MPH; Mollie Marr[23], BFA; Ajay Major[24], MD, MBA; Jason V Meisel[25], PMHNP-BC; Erika A Petersen[26], MD, FAANS, FACS; Barak Raguan[27], MD; Allison Rhodes[28], BA; Deborah D Rupert[29,30], MS, MA; Nadia A Sam-Agudu[31,32], MD, CTropMed; Naledi Saul[33], MPM; Jarna R Shah[26], MD; Lisa Kennedy Sheldon[34], PhD, ANP-BC, AOCNP, FAAN; Christian T Sinclair[35], MD; Kerry Spencer[36], PhD; Natalie H Strand[37], MD; Carl G Streed Jr[38], MD, MPH, FACP; Avery M Trudell[39], BSA

[1]School of Nursing, University of Virginia, Charlottesville, VA, United States
[2]Northeast Georgia Medical Center, Gainesville, GA, United States
[3]School of Medicine, Stanford University, Stanford, CA, United States
[4]Center for Public Health and Human Rights, Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD, United States
[5]College of Nursing, Medical University of South Carolina, Charleston, SC, United States
[6]School of Medicine, Washington University, St. Louis, MO, United States
[7]Department of General Internal Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
[8]Oncology Nursing Consultant, Newark, DE, United States
[9]Department of Pediatrics, University of Connecticut, Farmington, CT, United States
[10]Department of Respiratory Care, Boise State University, Boise, ID, United States
[11]Hematology/Oncology Section, Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States
[12]Swansea University Medical School, Swansea, United Kingdom
[13]Trent/Fleming School of Nursing, Peterborough, ON, Canada
[14]Department of Biomedical Sciences, University of South Carolina School of Medicine Greenville, Greenville, SC, United States
[15]College of Pharmacy, Glendale Campus, Midwestern University, Glendale, AZ, United States
[16]Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, United States
[17]Department of Pediatrics, Baylor College of Medicine, Houston, TX, United States
[18]Department of Neurology and Neurosurgery, Tulane University, New Orleans, LA, United States
[19]Division of Hematology and Oncology, University of Illinois Chicago, Chicago, IL, United States
[20]College of Nursing, Ohio State University, Columbus, OH, United States
[21]Department of Neurology, University of California Los Angeles, Los Angeles, CA, United States
[22]School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
[23]School of Medicine, Oregon Health & Science University, Portland, OR, United States
[24]Section of Hematology and Oncology, University of Chicago, Chicago, IL, United States
[25]Hunter School of Nursing, City University of New York, New York, NY, United States
[26]University of Arkansas for Medical Sciences, Little Rock, AR, United States
[27]Meir Medical Center, Kfar Saba, Israel
[28]School of Medicine, Tufts University, Boston, MA, United States
[29]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, United States
[30]State of New York-Stony Brook University, Stony Brook, NY, United States
[31]Institute of Human Virology and Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, United States
[32]International Research Center of Excellence, Institute of Human Virology Nigeria, Abuja, Nigeria

[33]Office of Career and Professional Development, University of California San Francisco, San Francisco, CA, United States
[34]Oncology Nursing Society, Pittsburgh, PA, United States
[35]University of Kansas Health System, Kansas City, KS, United States
[36]Department of Mathematics and Physics, Stevenson University, Owings Mills, MD, United States
[37]Department of Anesthesiology, Mayo Clinic, Phoenix, AZ, United States
[38]Department of Medicine, School of Medicine, Boston University, Boston, MA, United States
[39]McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, United States

Corresponding Author:
Kimberly D Acquaviva, PhD, MSW
School of Nursing
University of Virginia
4005 McLeod Hall
Charlottesville, VA, 22903
United States
Phone: 1 202 423 0984
Email: kda8xj@virginia.edu

Related Articles:

This is a corrected version. See correction statement: https://www.jmir.org/2020/12/e26225/

Comment in: https://www.jmir.org/2021/6/e27664

Abstract

Background: The traditional model of promotion and tenure in the health professions relies heavily on formal scholarship through teaching, research, and service. Institutions consider how much weight to give activities in each of these areas and determine a threshold for advancement. With the emergence of social media, scholars can engage wider audiences in creative ways and have a broader impact. Conventional metrics like the h-index do not account for social media impact. Social media engagement is poorly represented in most curricula vitae (CV) and therefore is undervalued in promotion and tenure reviews.

Objective: The objective was to develop crowdsourced guidelines for documenting social media scholarship. These guidelines aimed to provide a structure for documenting a scholar's general impact on social media, as well as methods of documenting individual social media contributions exemplifying innovation, education, mentorship, advocacy, and dissemination.

Methods: To create unifying guidelines, we created a crowdsourced process that capitalized on the strengths of social media and generated a case example of successful use of the medium for academic collaboration. The primary author created a draft of the guidelines and then sought input from users on Twitter via a publicly accessible Google Document. There was no limitation on who could provide input and the work was done in a democratic, collaborative fashion. Contributors edited the draft over a period of 1 week (September 12-18, 2020). The primary and secondary authors then revised the draft to make it more concise. The guidelines and manuscript were then distributed to the contributors for edits and adopted by the group. All contributors were given the opportunity to serve as coauthors on the publication and were told upfront that authorship would depend on whether they were able to document the ways in which they met the 4 International Committee of Medical Journal Editors authorship criteria.

Results: We developed 2 sets of guidelines: Guidelines for Listing All Social Media Scholarship Under Public Scholarship (in Research/Scholarship Section of CV) and Guidelines for Listing Social Media Scholarship Under Research, Teaching, and Service Sections of CV. Institutions can choose which set fits their existing CV format.

Conclusions: With more uniformity, scholars can better represent the full scope and impact of their work. These guidelines are not intended to dictate how individual institutions should weigh social media contributions within promotion and tenure cases. Instead, by providing an initial set of guidelines, we hope to provide scholars and their institutions with a common format and language to document social media scholarship.

(J Med Internet Res 2020;22(12):e25070) doi:10.2196/25070

KEYWORDS

social media; promotion; tenure; health professions; scholarship; medicine; research; accomplishment; crowdsource; contribution; innovation; education; dissemination

Introduction

Background

Traditionally, promotion and tenure committees have focused on evaluating a candidate scholar's curriculum vitae (CV), educational dossier, or portfolio to determine their academic productivity and impact. When a scholar is thought to have reached a certain threshold of productivity and impact, they are rewarded with promotion from Assistant to Associate Professor, or from Associate to full Professor. Individual institutions have always had leeway in determining thresholds for advancement, and there is wide interinstitutional variability in this determination [1]. Informally, candidates are considered ready for advancement to Associate Professor when they develop a national reputation and advancement to full Professor when they develop an international reputation. Within health professions schools (schools of medicine, nursing, public health, etc), scholarly accomplishment for the sake of promotion and tenure has typically been divided into 3 categories: teaching (eg, traditional classroom teaching, teaching lectures, invited lectures); research (eg, bench research, clinical trials, surveys, case series); and service (eg, committees, professional organization leadership, community advocacy, journal reviews). Traditionally, an individual scholar has documented productivity in research, teaching, and service by listing lectures given in certain venues, abstracts accepted at conferences, committees served on, or papers published in scientific journals. Activities within each category might have different levels of impact (ie, different weightings for the promotion and tenure committee). For example, a teaching lecture is usually considered less impactful than an invited lecture, and a grand rounds lecture at a local hospital is generally considered to be less impactful than a lecture at a national convention.

A new model of evaluating and documenting academic accomplishments was last introduced into promotion and tenure criteria at health professions schools in the United States in 1990, when Boyer [2] argued that scholarship should be reconceptualized from encompassing task-based silos to categories of discovery, integration, application, and teaching. Some schools integrated the Boyer model into their promotion and tenure review criteria, whereas others retained the traditional research, teaching, and service model. However, Boyer's categories remained closely aligned with the task-based silos the model sought to reject, rebranding the task-based silos rather than transcending them. In this paper we refer to the categories of research, teaching, and service with the understanding that some institutions may use Boyer's terminology instead.

Social media platforms (eg, Twitter, Facebook, Reddit, Instagram, and TikTok) now provide scholars with a significant platform for engaging in and disseminating scholarship. For example, Twitter, a social media platform with robust health researcher engagement [3,4], has over 186 million active daily users [5]. Instagram has over 1 billion active daily users [6], and TikTok has 800 million active users [7]. Each of these social media platforms has greater academic and nonacademic reach than all of the medical and nursing journals in the world combined, but none of the health professions has a widely accepted model for scholars to document their social media engagement on CVs or in promotion dossiers.

While academic engagement on social media is not free from entanglement—including lack of traditional peer review, potential for disinformation, and intermixing of nonacademic content—there remains a tremendous potential for teaching, dissemination of research, and service. Conventional metrics of academic productivity, including bibliometric indices, such as the h-index, which are now used by some institutions in promotions decisions [8-10], do not account for these new modes of scholarly engagement and dissemination. Over the past three decades, as academic engagement has become increasingly virtual, institutions have struggled to determine how to recognize and measure academic impact. How does hosting a podcast compare to teaching a semester-long course? How does publishing a page (eg, a blog post) on a highly visited website compare to publishing in an academic journal? How does advocacy on a social media site compare to advocacy within a hospital committee? Do newer metrics, such as altmetrics, that incorporate social media engagement for academic content predict future research productivity [11,12]?

Existing Guidelines for Including Social Media in the Promotion and Tenure Process

Some institutions such as the Mayo Clinic in the United States have led the effort to recognize the scholarly impact of social media engagement and are already considering these contributions in the promotion and tenure process [13]. Similarly, the US Council of Emergency Medicine Residency Directors proposed a formal set of guidelines for including digital scholarship in this process [14]. O'Glasser et al [15] called for the recognition of social media engagement earlier in health professional careers, and provided guidance for scholars-in-training on how best to include creative and social media work in standardized applications such as the Electronic Residency Application Service for medical residencies and fellowships. Cabrera et al [12,16] and Gruzd et al [17] outlined potential categories for documentation of social media activities in the CV. Sherbino et al [18] described a consensus process for determining criteria for social media–based scholarship to inform which activities can and may be listed on a CV. Their consensus was that scholarship should be based on 4 criteria. First, the topic must be original, stemming from the authors themselves. Second, the topic should advance the field of health education through innovative theories, research, or best practices. Third, scholarship should be able to be saved, archived, and easily disseminated to the masses. Fourth, scholarship should provide an avenue for the public to comment and provide feedback in a way that elicits further discussion. Additionally, using a modified Delphi methodology to survey experts among the scientific community, Lin et al [19] identified, via an at least 90% consensus, 13 quality indicators classified into 3 domains, credibility (n=8), content (n=4), and design (n=1), that provide the groundwork to judge future social media scholarship. With a continued rise in social media use, future research should develop a method for stakeholders to structure and stratify criteria for their individual purpose. These guidelines were echoed by Shapiro et al [20], who also included recommendations related to formatting, such as including hyperlinks when relevant. Finally, Cabrera et al [16] suggested the creation of a social media portfolio similar to an education portfolio to enhance the material included in the CV.

Thus, various and sometimes divergent guidelines exist for including social media in the promotion and tenure process, but there is currently no unifying structure.

Objective

Guidelines that are followed across institutions and are able to capture social media engagement in a meaningful way would simplify and unify communication of academic social media engagement. While there is a great deal of interinstitutional variability in how individual activities are valued, there is a pressing need for the widespread adoption of a method for documenting social media engagement in the academic CVs and dossiers of scholars in the health professions. We propose and describe a method for scholars in the health professions to cite social media activities and impact.

Before we provide these guidelines, however, we wish to issue one important caveat: scholars should neither be expected nor required to be active on social media in order to earn promotion or tenure. There are numerous reasons why someone may choose not to have a social media presence, and no scholar should be expected to disclose their reasons for making that decision. The guidelines provided in this paper should be adopted by institutions as an option for scholars to incorporate into their CV, not as a required section that must be included.

Methods

Framework for Open Collaboration

To create a set of unifying guidelines, we created a crowdsourced process that capitalized on the strengths of social media and generated a case example of successful use of such a medium for productive academic collaboration. In constructing a model for describing academic social media impact on a CV and in a dossier, we considered the following topics: (1) how to represent overall social media impact, (2) how to represent specific social media contributions, (3) where on the CV or dossier to list social media contributions, (4) the appropriate format for listing these contributions, and (5) how to demonstrate the reach or impact of a specific contribution.

The first author (KDA)—a tenured professor who has substantial experience writing, revising, and applying promotion and tenure criteria over the course of her career—created an initial draft of the guidelines after conducting a brief search of the literature to identify previous work describing how to cite social media impact on a CV or in the context of a promotion and tenure committee review. She then sought input from users on Twitter via a publicly accessible Google Document. There was no screening process for who could or could not join in the collaboration. Anyone who wanted to collaborate was welcome, regardless of discipline, specialty, title, country of residence, or degree completion. All contributors were given the opportunity to serve as coauthors on the publication and were told upfront that authorship would depend on whether they were able to document the ways in which they met the 4 International Committee of Medical Journal Editors (ICMJE) authorship criteria [21].

Crowdsourced Collaboration and Consensus

Contributors edited the draft over a period of 1 week (September 12-18, 2020). The primary and secondary authors then revised the draft to make it more concise. The guidelines and manuscript were then distributed to the list of contributors for edits and adopted by the group. When including specific examples or in discussing formatting, we used Twitter as the representative social media platform because of its heavy academic health presence, and also because it was the medium used to create these recommendations. These Twitter examples should be generally applicable to other social media platforms. Content not used in the primary article was extracted and used to form the basis of a second article (unpublished data) on the benefits of social media engagement for health professionals in academia. The second article went through a similar process for editing and shared authorship.

Results

Proposed Model for Tracking Social Media Impact

Impact of Social Media

One of the primary difficulties in describing the scholarly impact of social media is that social media users and accounts do not exist in a purely academic or scholarly space, but rather encompass a spectrum between the academy, the personal, and the wider society. A social media account that highlights the latest medical studies and attempts to provide educational content, for example, may also comment on current events or post personal pictures of family or pets. Simply describing an individual’s overall social media impact may place too much impact on popularity and broad appeal rather than true scholarly impact. On the other hand, focusing on individual and specific posts ignores the impact of being able to disseminate scholarly content to a wider audience through a broad social media following [22]. That being said, public scholars benefit from building a relationship with the general public through their posts, including those that are unrelated to scholarly work. Isaacson and Looman [22] noted that social media offers a unique opportunity to provide a bridge between and across multiple networks of both laypeople and professionals, which ultimately builds social capital. It is important that promotion and tenure reviewers do not devalue a scholar’s seemingly casual interactions on social media. Establishing an authentic social media presence (“branding”) serves to further one’s reach and the subsequent engagement and dissemination of ideas.

Therefore, we propose the following guidelines for documenting scholarly impact through social media in the CV and dossier for the sake of promotion and tenure, and which attempt to incorporate the importance of overall impact and popularity, as well as the impact of individual social media posts.

Documenting Overall Social Media Impact on the CV

If an institution has a CV section for public scholarship (or similar), a section for social media scholarship should be inserted here. If no institutional section requirement exists, then social media scholarship may be included as a separate high-level section within the research or scholarship section. For each social media platform on which an individual is active, they should create a separate entry that includes the following: username, dates active, number of followers, and platform-specific metrics regarding overall reach (eg, total impressions on Twitter, engagement rate for YouTube channel). Scholars may choose to list whether their social media account is “verified” but, given the nontransparency of the verification process, institutions should not give significant weight to verification status.

Documenting Specific Social Media Contributions on the CV

High-quality and high-impact contributions that are scholarly in nature or that promote the individual scholar’s academic mission should be cited in a CV similar to citing individual publications, abstracts, lectures, or courses taught. The location for citing social media contributions will depend on the requirements of the individual institution. Some institutions require scholarly output to be divided into the traditional research, teaching, and service categories. Other institutions might divide categories into the Boyer categories or have other categories such as advocacy, dissemination, or mentorship. In this article we provide 2 different sets of guidelines that can be inserted into an institution’s dossier and CV format depending on institutional preferences.

No set threshold exists for when a scholar should include an individual social media contribution on their CV and when to exclude it. In other words, when is a social media post impactful enough to be included on one’s CV? Since scholars move between institutions over time, institutions should not attempt to define what an appropriate threshold is for listing an individual social media contribution on a scholar’s CV. Instead, the onus should be on scholars to provide a brief explanation of why they chose to include a particular social media contribution on their CV.

Documenting Social Media Impact in an Academic Dossier

The purpose of the academic dossier is to provide a promotion and tenure review committee with representative examples of an individual’s scholarly output and to discuss the impact of this output and how it supports their academic mission. For example, a dossier might include representative slides from a scholar’s national presentation with a discussion of why it was impactful, how many times it has been given in various venues, and how many times it has been cited by other presentations. Or a dossier might include a written curriculum for a course that a scholar taught, including a discussion about course logistics, why the approach to the material was innovative or impactful, and how many learners the course impacted. Or, as a last example, a dossier might include an example of an individual scholar’s important publications, including the number of downloads or views of that publication, awards it won, or number of times it was cited in other publications.

Similarly, scholars in the health professions should include highly impactful and representative social media contributions in their dossier. These could include contributions from any of the categories described above but should be particularly meaningful or impactful. When including social media contributions in a dossier, the scholar should include links to each contribution, screenshots, advanced metrics indicating impact (number of likes, replies, retweets, and views), and an explanation of why the scholar chose to highlight each of these contributions. In the narrative section of the dossier, the scholar should consider including a discussion of the impact of the selected social media contributions, including impact on the scholar’s future scholarly output, how the post impacted other scholars or learners, and how and when the post was cited by other posts or in other media.

To standardize the way that scholars document and present social media contributions in their CVs and dossiers, we developed 2 sets of guidelines: Guidelines for Listing All Social Media Scholarship Under Public Scholarship (in Research/Scholarship Section of CV) and Guidelines for Listing Social Media Scholarship Under Research, Teaching, and Service Sections of CV. The content of both sets of guidelines is identical, and institutions can choose which set fits their existing CV format.

Guidelines for Listing All Social Media Scholarship Under Public Scholarship (in Research/Scholarship Section of CV)

Textbox 1 shows the first set of guidelines.

Textbox 1. Guidelines for Listing All Social Media Scholarship Under Public Scholarship (in Research/Scholarship Section of the Curriculum Vitae).

Public Scholarship - *Social Media Scholarship*

Overall Reach (Time Period: x/x/xx to y/y/xx)

[Platform]: [username]

Number of Followers/Subscribers/Connections:

Number of [Tweets, Posts, Videos, etc]:

Total Impressions and/or Other Platform-Specific Metrics:

Select Social Media Contributions

Innovation (contributions that propose new ideas)

- Link to contributions:
- Number of impressions:
- Explanation of why the scholar chose to highlight this:

Dissemination (contributions that share resources and/or findings)

- Link to contributions:
- Number of impressions:
- Explanation of why the scholar chose to highlight this:

Education (contributions that teach people something)

- Link to contribution:
- Number of impressions:
- Explanation of why the scholar chose to highlight this:

Advocacy (contributions about changing laws, policies, practices, and/or systems)

- Link to contribution:
- Number of impressions:
- Explanation of why the scholar chose to highlight this:

Mentorship (contributions about mentees achievements)

- Link to contribution:
- Number of impressions:
- Explanation of why the scholar chose to highlight this contribution:
- Presentations/Chats/Blogs/Podcasts Delivered Via Social Media

[Platform] Chats

- Name of [Platform] Chat:
- Role: [host, cohost, etc]
- If longitudinal, list time period and cite metrics. If discrete, list dates and topics:
- Explanation of why the scholar chose to highlight this contribution:

[Platform] Live Video

- Host or cohost(s) of live feed:
- Individual sessions [Date(s), Topic(s), Engagement/Reach]:
- Permanent Link:
- Explanation of why the scholar chose to highlight this contribution:

[Platform] Recorded Video

- Host or cohost(s) of recorded video:
- Title of Video:
- Date(s), Topic(s), Engagement/Reach:
- Permanent Link:

Explanation of why the scholar chose to highlight this contribution:

Blog Posts

Author:

Title of Blog:

Organization/Entity Publishing the Blog:

Date Published:

Link to Blog:

Explanation of why the scholar chose to highlight this contribution:

Podcasts

Role: [Host or Guest]

Name of Podcast:

Date Released:

Link to Podcast Episode:

Explanation of why the scholar chose to highlight this contribution:

Infographics/Other Visuals

Infographic

Title of Infographic:

Date Posted/Published:

Link to Infographic:

Explanation of why the scholar chose to highlight this contribution:

Other Visuals

Author/Creator:

Title of Visual:

Date Posted/Published:

Link to Visual:

Explanation of why the scholar chose to highlight this contribution:

Guidelines for Listing Social Media Scholarship Under Research, Teaching, and Service Sections of CV

Textbox 2 shows the second set of guidelines.

Textbox 2. Guidelines for Listing Social Media Scholarship Under Research, Teaching, and Service Sections of the Curriculum Vitae.

RESEARCH - *Social Media Scholarship*

Overall Reach (Time Period: x/x/xx to y/y/xx)

[Platform]: [username]

Number of Followers/Subscribers/Connections:

Number of [Tweets, Posts, Videos, etc]:

Total Impressions and/or Other Platform-Specific Metrics:

Select Social Media Contributions

Innovation (contributions that propose new ideas)

- Link to contributions:
- Number of impressions:
- Explanation of why the scholar chose to highlight this:

Dissemination (contributions that share resources and/or findings)

- Link to contributions:
- Number of impressions:
- Explanation of why the scholar chose to highlight this
- Scholarly Presentations/Chats/Blogs/Podcasts Delivered Via Social Media

[Platform] Chats

- Name of [Platform] Chat:
- Role: [host, cohost, etc]
- If longitudinal, list time period and cite metrics:
- If discrete, list dates and topics:
- Explanation of why the scholar chose to highlight this contribution:

[Platform] Live Video

- Host or cohost(s) of live feed:
- Individual sessions [Date(s), Topic(s), Engagement/Reach]:
- Permanent Link:
- Explanation of why the scholar chose to highlight this contribution:

[Platform] Recorded Video

- Host or cohost(s) of recorded video:
- Title of Video:
- Date(s), Topic(s), Engagement/Reach:
- Permanent Link:
- Explanation of why the scholar chose to highlight this contribution:

Blog Posts

- Author:
- Title of Blog:
- Organization/Entity Publishing the Blog:
- Date Published:
- Link to Blog:
- Explanation of why the scholar chose to highlight this contribution:

Podcasts

- Role: [Host or Guest]
- Name of Podcast:

Date Released:

Link to Podcast Episode:

Explanation of why the scholar chose to highlight this contribution:

Scholarly Infographics/Other Visuals

Infographic

Title of Infographic:

Date Posted/Published:

Link to Infographic:

Explanation of why the scholar chose to highlight this contribution:

Other Visuals

Author/Creator:

Title of Visual:

Date Posted/Published:

Link to Visual:

Explanation of why the scholar chose to highlight this contribution:

TEACHING

Select Social Media Contributions

Education (contributions that teach people something)

Link to contribution:

Number of impressions:

Explanation of why the scholar chose to highlight this:

Educational Presentations/Chats/Blogs/Podcasts Delivered Via Social Media

[Platform] Chats

Name of [Platform] Chat:

Role: [host, cohost, etc]

If longitudinal, list time period and cite metrics. If discrete, list dates and topics:

Explanation of why the scholar chose to highlight this contribution:

[Platform] Live Video

Host or cohost(s) of live feed:

Individual sessions [Date(s), Topic(s), Engagement/Reach]:

Permanent Link:

Explanation of why the scholar chose to highlight this contribution:

[Platform] Recorded Video

Host or cohost(s) of recorded video:

Title of Video:

Date(s), Topic(s), Engagement/Reach:

Permanent Link:

Explanation of why the scholar chose to highlight this contribution:

Blog Posts

Author:

Title of Blog:

Organization/Entity Publishing the Blog:

Date Published:

Link to Blog:

Explanation of why the scholar chose to highlight this contribution:

Podcasts

Role: [Host or Guest]

Name of Podcast:

Date Released:

Link to Podcast Episode:

Explanation of why the scholar chose to highlight this contribution:

Educational Infographics/Other Visuals

Infographic

Title of Infographic:

Date Posted/Published:

Link to Infographic:

Explanation of why the scholar chose to highlight this contribution:

Other Visuals

Author/Creator:

Title of Visual:

Date Posted/Published:

Link to Visual:

Explanation of why the scholar chose to highlight this contribution:

SERVICE

Select Social Media Contributions

Advocacy (contributions about changing laws, policies, practices, and/or systems)

Link to contribution:

Number of impressions:

Explanation of why the scholar chose to highlight this:

Mentorship (contributions about mentees achievements)

Link to contribution:

Number of impressions:

Explanation of why the scholar chose to highlight this contribution.

Discussion

Limitations of the Guidelines

In developing these guidelines, we attempted to account for the breadth of social media platforms, as well the variability in institutional guidelines and expectations for promotion and tenure. Given the growing list of social media platforms and the wide variety of ways in which users can engage with them, developing a set of immutable guidelines was neither possible nor practical. Social media platforms will continue to evolve, and new platforms will continue to be introduced that provide novel ways for health scholars to interact with each other and the public, but will continue to raise new challenges about the scope and nature of academic impact on these platforms.

Current metrics for measuring the impact of social media use, whether an individual’s overall social media impact or the impact of a specific contribution or facet of a scholar’s social media presence, rely on blunt and nonspecific markers of impact, including numbers of followers, numbers of likes and shares, and other social media users’ amount of engagement. Our hope is that as research continues to show the general impact of social media on health scholarship, more refined metrics for measuring individual impacts will become apparent.

We recognize that some institutions remain unsure of the relative impact of social media scholarship in the academic realm. For the purpose of developing these guidelines, we specifically chose not to discuss the merits of social media in this paper, but rather to assume that scholarly impact within social media is worthwhile and worthy of inclusion in the promotion and tenure process.

Conclusion

This paper presents crowdsourced guidelines on how to cite scholarly productivity on social media, while recognizing the limitations inherent in measuring social media impact across a variety of academic institutions. These guidelines describe a process and structure for documenting and describing a scholar's general impact on a social media platform, as well as methods of documenting individual social media contributions such as teaching posts, mentorship, and ongoing advocacy.

Since promotion and tenure is decided by individual academic institutions based on the criteria developed and adopted by their faculty, these guidelines are not intended to dictate how individual institutions should weigh social media contributions within promotion and tenure cases. Instead, by providing an initial set of guidelines, we hope to provide scholars and their institutions with a common format and language to describe what is becoming more and more ubiquitous among academics.

We expect that as social media adoption and use continues to grow among academics, and as new social media platforms arise and new methods of applying social media to health education and scholarly distribution become apparent, research will continue to demonstrate the impact of social media on education, potentially affecting the scope of distribution, health care advocacy and equity in traditionally inequitable fields (such as gender and race), and scholarly innovation.

Acknowledgments

All authors met the ICMJE authorship criteria, including reviewing and approving this manuscript for publication. Each of the 40 coauthors completed an Authorship Attestation Form confirming that they met all 4 of the ICMJE authorship criteria. In addition, each coauthor described on a shared Google Sheet how they made "substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data." The authors wish to thank Jocelyn C Anderson, PhD, RN, S Alexander Kemery, PhD, RN, Rebecca Koszalinski PhD, RN, CRRN, CMSRN, and Chrystal L Lewis, PhD, RN for their comments on the draft of this manuscript.

Conflicts of Interest

EAP is a consultant for Nevro, Abbott Neuromodulation, and Medtronic, with research and educational support from Nevro, Neuros Medical, ReNeuron, Saluda, and SPR Therapeutics. EAP has stock options with SynerFuse. CTS owns less than US $5000 in stock shares of Twitter, and less than US $15,000 in stock shares of Alphabet (YouTube).

References

1. Atasoylu AA, Wright SM, Beasley BW, Cofrancesco J, Macpherson DS, Partridge T, et al. Promotion criteria for clinician-educators. J Gen Intern Med 2003 Sep;18(9):711-716 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.10425.x] [Medline: 12950479]
2. Boyer EL. Scholarship Reconsidered: Priorities of the Professoriate. Stanford, CA: Carnegie Foundation for the Advancement of Teaching; 1990.
3. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Dec;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
4. Diug B, Kendal E, Ilic D. Evaluating the use of Twitter as a tool to increase engagement in medical education. Educ Health (Abingdon) 2016;29(3):223-230 [FREE Full text] [doi: 10.4103/1357-6283.204216] [Medline: 28406107]
5. Goldsmith J. Twitter active daily users surge 34% to record 186M in Q2, revenue dips, CEO Jack Dorsey apologizes for breach. Deadline. 2020 Jul 23. URL: https://deadline.com/2020/07/twitter-active-daily-users-surge-34-to-186m-q2-revenue-dips-19-1202992835/ [accessed 2020-09-21]
6. Clement J. Instagram: age distribution of global audiences 2020. Statista. 2020 Jul 24. URL: https://www.statista.com/statistics/325587/instagram-global-age-group/ [accessed 2020-09-21]
7. Mohsin M. 10 TikTok statistics that you need to know. Oberlo. 2020 Jul 03. URL: https://www.oberlo.com/blog/tiktok-statistics [accessed 2020-09-21]
8. DeLuca LA, St John A, Stolz U, Matheson L, Simpson A, Denninghoff KR. The distribution of the h-index among academic emergency physicians in the United States. Acad Emerg Med 2013 Oct;20(10):997-1003 [FREE Full text] [doi: 10.1111/acem.12226] [Medline: 24127702]
9. Yuan JT, Aires DJ, DaCunha M, Funk K, Habashi-Daniel A, Moore SA, et al. The h-index for associate and full professors of dermatology in the United States: an epidemiologic study of scholastic production. Cutis 2017 Dec;100(6):395-398. [Medline: 29360887]
10. Tschudy MM, Rowe TL, Dover GJ, Cheng TL. Pediatric academic productivity: pediatric benchmarks for the h- and g-indices. J Pediatr 2016 Feb;169:272-276. [doi: 10.1016/j.jpeds.2015.10.030] [Medline: 26563533]
11. Maggio LA, Leroux TC, Meyer HS, Artino AR. #MedEd: exploring the relationship between altmetrics and traditional measures of dissemination in health professions education. Perspect Med Educ 2018 Aug;7(4):239-247 [FREE Full text] [doi: 10.1007/s40037-018-0438-5] [Medline: 29949099]
12. Cabrera D, Vartabedian BS, Spinner RJ, Jordan BL, Aase LA, Timimi FK. More than likes and tweets: creating social media portfolios for academic promotion and tenure. J Grad Med Educ 2017 Aug;9(4):421-425 [FREE Full text] [doi: 10.4300/JGME-D-17-00171.1] [Medline: 28824752]

13. Cabrera D. Mayo Clinic includes social media scholarship activities in academic advancement. Mayo Clinic Social Media Network. 2016 May 25. URL: https://socialmedia.mayoclinic.org/2016/05/25/mayo-clinic-includes-social-media-scholarship-activities-in-academic-advancement/ [accessed 2020-09-21]
14. Husain A, Repanshek Z, Singh M, Ankel F, Beck-Esmay J, Cabrera D, et al. Consensus guidelines for digital scholarship in academic promotion. West J Emerg Med 2020 Jul 08;21(4):883-891 [FREE Full text] [doi: 10.5811/westjem.2020.4.46441] [Medline: 32726260]
15. O'Glasser AY, Wray C, Aurora V, Shapiro M. White paper: social media, podcasts, and blogs on a residency/fellowship ERAS application.: Explore the Space, LLC; 2019 Jun. URL: https://www.explorethespaceshow.com/wp-content/uploads/2019/06/White-Paper-2-Final.pdf [accessed 2020-09-21]
16. Cabrera D, Roy D, Chisolm MS. Social media scholarship and alternative metrics for academic promotion and tenure. J Am Coll Radiol 2018 Jan;15(1 Pt B):135-141. [doi: 10.1016/j.jacr.2017.09.012] [Medline: 29122503]
17. Gruzd A, Staves K, Wilk A. Tenure and promotion in the age of online social media. Proc Am Soc Info Sci Tech 2012 Jan 11;48(1):1-9. [doi: 10.1002/meet.2011.14504801154] [Medline: 25855820]
18. Sherbino J, Arora VM, Van Melle E, Rogers R, Frank JR, Holmboe ES. Criteria for social media-based scholarship in health professions education. Postgrad Med J 2015 Oct;91(1080):551-555. [doi: 10.1136/postgradmedj-2015-133300] [Medline: 26275426]
19. Lin M, Thoma B, Trueger NS, Ankel F, Sherbino J, Chan T. Quality indicators for blogs and podcasts used in medical education: modified Delphi consensus recommendations by an international cohort of health professions educators. Postgrad Med J 2015 Oct;91(1080):546-550. [doi: 10.1136/postgradmedj-2014-133230] [Medline: 26275428]
20. Shapiro M, Wray C, Aurora V. White paper: social media, podcasts, and blogs on a professional curriculum vitae.: Explore the Space, LLC; 2019 Apr 02. URL: https://www.explorethespaceshow.com/podcasting/white-paper-on-social-media-cvs/ [accessed 2020-09-21]
21. nternational Committee of Medical Journal Editors. Defining the role of authors and contributors.: ICMJE; 2020. URL: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html [accessed 2020-09-24]
22. Isaacson K, Looman WS. Strategies for developing family nursing communities of practice through social media. J Fam Nurs 2017 Feb;23(1):73-89. [doi: 10.1177/1074840716689078] [Medline: 28795934]

Abbreviations

CV: curriculum vitae
ICMJE: International Committee of Medical Journal Editors

Edited by G Eysenbach; submitted 16.10.20; peer-reviewed by A Alasmari, A Narang, J M. Faieta; comments to author 06.11.20; revised version received 11.11.20; accepted 16.11.20; published 02.12.20.

Please cite as:
Acquaviva KD, Mugele J, Abadilla N, Adamson T, Bernstein SL, Bhayani RK, Büchi AE, Burbage D, Carroll CL, Davis SP, Dhawan N, Eaton A, English K, Grier JT, Gurney MK, Hahn ES, Haq H, Huang B, Jain S, Jun J, Kerr WT, Keyes T, Kirby AR, Leary M, Marr M, Major A, Meisel JV, Petersen EA, Raguan B, Rhodes A, Rupert DD, Sam-Agudu NA, Saul N, Shah JR, Sheldon LK, Sinclair CT, Spencer K, Strand NH, Streed Jr CG, Trudell AM
Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions
J Med Internet Res 2020;22(12):e25070
URL: https://www.jmir.org/2020/12/e25070
doi:10.2196/25070
PMID:33263554

Original Paper

Social Media Use for Health Communication by the CDC in Mainland China: National Survey Study 2009-2020

Runxi Zeng[1], PhD; Menghan Li[1], MD

Center for Communication and Social Development, School of Journalism and Communication, Chongqing University, Chongqing, China

Corresponding Author:
Runxi Zeng, PhD
Center for Communication and Social Development
School of Journalism and Communication
Chongqing University
South Road 55, Daxuecheng, Shapingba District
Chongqing, 401331
China
Phone: 86 18971570595
Email: zrx@cqu.edu.cn

Abstract

Background: In recent years, public health incidents that pose a serious threat to public life have occurred frequently in China. The use of social media by public health authorities has helped to reduce these threats by increasing effective risk communication between the government and the public.

Objective: The aim of this study is to reveal how China's Center for Disease Control and Prevention (CDC) uses social media to improve three aspects of health communication between the government and the public: adoption, operation, and interaction.

Methods: To analyze the 134 CDC government Weibo accounts at the provincial- and prefecture-level administration regions in mainland China, we collected their account data and extracted 1215 Weibo tweets. We also supplemented the data to reveal the overall performance of the CDC's government Weibo use during the COVID-19 crisis.

Results: The registration rate of the CDC's government Weibo accounts increased year by year, and the local authorities registered Weibo accounts before the central government authorities. In total, 29.8% (n=134) of the 450 CDC facilities have registered an account. Among the 134 CDC facilities that have registered Weibo accounts, the registration rate in the eastern region (n=68, 50.7%) was higher than those in the central region (n=30, 22.4%) and the western region (n=36, 26.9%). Nearly 90.0% of these Weibo accounts had official certification, but there were dropouts in the specific operating process. One-third of the accounts have not been updated for more than 1 year, and the number of Weibo followers was polarized, with a maximum and minimum difference of 1 million. The response rate to users' comments was less than 1%. Emergency information, multimedia content, and original content were more helpful in promoting communication between the government and the public. Such interaction was partially improved during the COVID-19 pandemic. The CDC updated the daily epidemic situation and provided popular science information for epidemic prevention and control for the public in a timely manner.

Conclusions: China's CDC is using more social media to popularize daily health information and has taken the first step to improve communication between the government and the public. However, equal dialogue, two-way interactions, and effective communication with the public still need improvement.

(J Med Internet Res 2020;22(12):e19470) doi:10.2196/19470

KEYWORDS

social media; public health agencies; Center for Disease Control and Prevention; China; government Weibo; COVID-19

Introduction

One of the keys to dealing with public health emergencies is timely and effective risk communication. Through a dialogue between the government and the public, we can minimize the information asymmetry between the government and the public, and help the public to take preventive measures quickly [1]. With the development of information and communication technology, social media, with the characteristics of participation, openness, dialogue, and other tools, has brought unprecedented opportunities for improving communication between the government and the public [2,3]. Thus, social media can improve the ability of government management to strengthen communication between the government and the public, and to realize an open government [4]. For example, previous studies have shown that the use of social media by the government can promote cross-sectoral information sharing [5], increase government transparency [6], promote public political participation [7], and help the public to respond to disasters [8].

Public health authorities are also increasingly using social media for information disclosure and risk communication. Previous studies have analyzed the main factors of public health authorities' willingness to adopt social media, such as organizational size and geographical location [9,10]. Studies have also shown that traditional media is used by public health authorities to respond to general health problems, while social media is used to respond to public health emergencies [11], and this role of responding to public health emergencies has been confirmed by other studies [12-14]. For example, analyzing the number of tweets and the degree of public participation can effectively predict the actual dynamics of an epidemic [15]. In addition, existing research also discusses the role of social media use by public health authorities in advancing health reform and building an open government [16,17]. In general, existing studies have explored the influencing factors, preferences, and significance of social media use, but they have paid less attention to how public health authorities use social media to communicate with the public. In particular, these studies have rarely investigated and evaluated the effects of communication [18].

Moreover, in the Western context, China is regarded as an authoritarian country that differs from the Western-style democratic system. Sudden public health incidents occur frequently in China and have an impact on China's economic development and social stability. Previous studies have attributed these effects to China's administrative system. These studies have noted that China's emergency decision making is often guided by a top-down command and control system, and that information transmission follows a layer-by-layer linear model. As a result, the public feedback channel is not smooth, and the interaction between the government and the public is limited [19,20]. For example, during the severe acute respiratory syndrome incident, there were delays in decision making caused by poor communication, which led to the spread of public panic and threatened public health and safety [19]. This situation has also occurred in the fight against COVID-19. Therefore, identifying ways to improve communication between the government and the public, and promoting timely and effective risk communication is key for China to deal with sudden public health incidents. However, there are few studies that have examined the current situation and the interaction of social media use by Chinese public health authorities from an empirical perspective.

The purpose of this study is to examine how Chinese public health authorities use social media to improve communication between the government and the public. This study analyzes the government affairs Weibo information of the Center for Disease Control and Prevention (CDC) at the provincial and prefectural levels in mainland China; describes the current situation of the adoption, operation, and communication of government Weibo accounts; discusses whether the use of social media by Chinese public health authorities has improved health communication between the government and the public; and discusses what factors help to promote communication between the government and the public. Specifically, this study aims to answer the following questions:

1. Is it common for the CDC at the provincial and prefectural levels in mainland China to use government Weibo accounts?
2. What is the current status of government Weibo use by the CDC at the provincial and prefectural levels in mainland China?
3. How does the CDC at the provincial and prefectural levels in mainland China communicate with the public on government Weibo accounts?

Methods

Study Sample

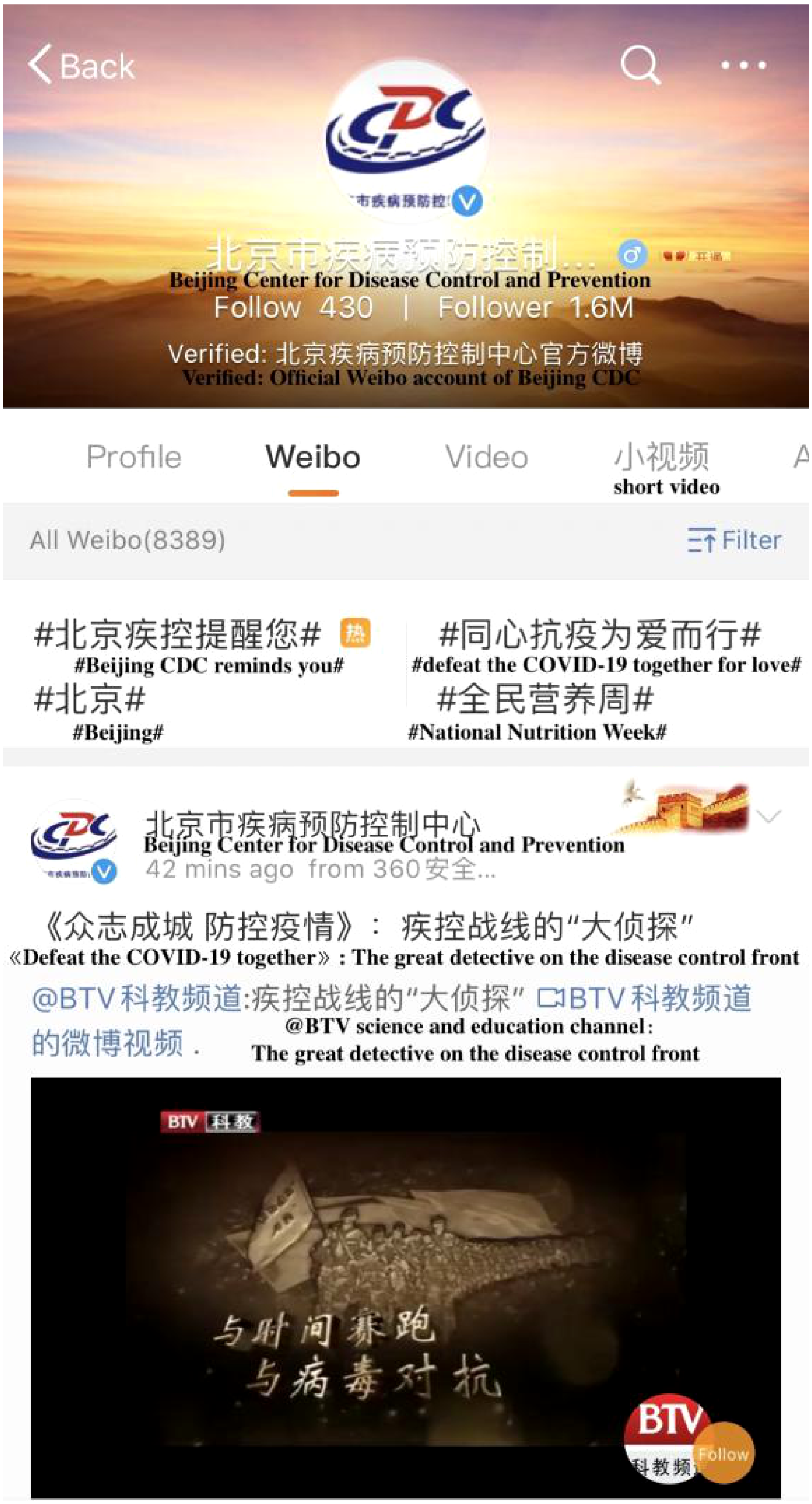
The CDC is a state-funded bureau under the leadership of the National Health Commission of China that specializes in disease control and prevention, and public health. According to the official website, its mission is to create a safe and healthy environment, maintain social stability, ensure national security and promote the health of people through the prevention and control of diseases, injuries, and disabilities. There are corresponding institutions from the central government to the local government. Sina Weibo is one of the most popular social media sites in China and is similar to Twitter. As of December 2019, there were 139,000 government agencies with registered Weibo accounts on the platform, according to the China Internet Network Information Center 45th Statistical Report on the Development of the Internet in China [21]. Because Weibo is highly influential, government agencies set up government Weibo accounts on the Sina Weibo platform.

A total of 134 sample accounts were collected. The sample collection process was as follows. First, through the "find-search-user" function on the Sina Weibo client, we conducted searches with "xx province + CDC" ("xx省+疾病预防控制中心"), "xx City + CDC" ("xx市+疾病预防控制中心"), "Centers for Disease Control and Prevention" ("疾控中心"), and "disease control and prevention" ("疾控") as the keywords for the search. For example, "Jiangsu Centers for Disease Control and Prevention" ("江苏省疾病预防控制中心"), "Nanjing Centers for Disease Control and Prevention" ("南京市疾病预防控制中心"), and "Chongqing Disease Control and Prevention" ("重庆疾控") are Weibo accounts that meet the requirements. Second, we visited the CDC's official website and used the Weibo account announced by the CDC on its official website as a sample. The samples were collected from March 21, 2019, to March 30, 2019. To understand the use of the CDC's government Weibo accounts during COVID-19, this study also observed the data for these 134 sample accounts from January 1 to June 30, 2020. To ensure that there were no omissions, the collection process was jointly undertaken by a teacher and two trained master's degree students, and the collection results of the three were found to be completely consistent.

Data Collection

Data collection for this study was mainly done using the Sina Weibo webpage. Figure 1 shows the Sina Weibo webpage of a CDC facility in China. We collected the latest 10 Weibo tweets posted by these 134 CDC government Weibo accounts before midnight on March 30, 2019. Considering that some Weibo accounts had posted less than 10 tweets or no tweets since their registration, we could not collect all 1340 tweets. Therefore, this study collected a total of 1215 Weibo tweets. The data collected mainly included topic type, content form, degree of originality, reply rate of comments, and reply time of comments. In addition, this study collected supplemental relevant data on the overall performances of the CDC's government Weibo accounts in the pandemic from January 1 to June 30, 2020.

Figure 1. Sina Weibo webpage of a Center for Disease Control and Prevention facility in China.



北京市疾病预防控制...
Beijing Center for Disease Control and Prevention
Follow 430 | Follower 1.6M
Verified: 北京疾病预防控制中心官方微博
Verified: Official Weibo account of Beijing CDC

#北京疾控提醒您#
#Beijing CDC reminds you#
#同心抗疫为爱而行#
#defeat the COVID-19 together for love#
#北京#
#Beijing#
#全民营养周#
#National Nutrition Week#

北京市疾病预防控制中心
Beijing Center for Disease Control and Prevention
42 mins ago from 360安全...

《众志成城 防控疫情》： 疾控战线的“大侦探”
《Defeat the COVID-19 together》: The great detective on the disease control front
@BTV科教频道:疾控战线的“大侦探” BTV科教频道的微博视频 .
@BTV science and education channel: The great detective on the disease control front

In this study, the adoption and operation of the CDC's government Weibo accounts were also included in the survey. In the survey on the adoption of government Weibo accounts, the account registration time was collected. In the survey on the operation of government Weibo accounts, the data collected included the following: whether they received official certification from Sina Weibo, the update time for the most recent Weibo post, and the number of followers.

Data Analysis

As China's regional economy has shown uneven development in the eastern, central, and western regions [22], which may have an impact on the development of the CDC's government Weibo accounts, this study is based on the division of the eastern, central, and western regions of mainland China by the National Bureau of Statistics. The CDC was divided into the "Eastern CDC," "Central CDC," and "Western CDC" for observation. In terms of data coding, first, whether the Weibo account passed the official certification of Sina Weibo was coded as passed or failed. Second, the year of registration of the CDC's government Weibo accounts was coded as follows: 2009-2011, 2012-2014, 2015-2017, or 2018-2019. Third, the update time of the most recent Weibo post was coded as follows: within 30 days, 31-90 days, 91-365 days, more than 365 days, or no content. In the analysis of the government Weibo accounts' interactions, this study used the content analysis method to code and analyze Weibo materials, as shown in Textbox 1. Before the formal coding, we analyzed the reliability of the three coders. The overall reliability was 0.97, and the lowest item reliability was 0.93, which met the reliability requirements of content analysis.

Textbox 1. Dimensions and indicators of 1215 Center for Disease Control and Prevention government Weibo tweets.

Topic type

- Disease control information
- Emergency information
- Popularization of health knowledge
- Popularization of disease knowledge
- Radiation hygiene/school hygiene
- Government affairs trends
- Policy interpretation
- Weibo help/citizen consultation
- Other

Content form

- Only original text
- Posts with "pictures/videos/hyperlinks"

Degree of originality

- Original posts
- Retweeted posts

Reply rate of comments

- Reply only once
- Interactive reply

Reply time of comments

- Between 0 hours and 1 hour
- Between 1 hour and 8 hours
- Between 8 and 12 hours
- More than 12 hours

Availability of Data and Materials

All the data are publicly available on the internet via the search strategy indicated in the Study Sample section. The original data are in Chinese and can be provided upon request.

Ethics Approval and Consent to Participate

The study was reviewed and approved by the Academic Committee of the School of Journalism and Communication at Chongqing University, which acts as an ethics committee. According to the committee's review report, the sample of this study are nonparticipants. Therefore, this study does not violate research ethics.

Results

Popularization of the CDC's Government Weibo Accounts

Geographical Distribution of the CDC's Government Weibo Adoption

There are a total of 450 CDC facilities in 31 provincial- and prefecture-level administrative regions in mainland China, as shown in Figure 2. In total, 31 provincial-level CDC government Weibo accounts should be registered, but only 8 have been registered, with a registration rate of 25.8%. Additionally, 419 prefectural-level CDC government Weibo accounts should be registered but only 126 have been registered, with a registration rate of 30.1%.

Figure 2. Distribution of the Center for Disease Control and Prevention facilities in mainland China.

There are regional differences in the adoption of the CDC's government Weibo accounts, and the registration rate shows a decreasing trend in the eastern region, the central region, and the western region. Table 1 provides the distribution of CDC facilities and the CDC's government Weibo accounts in mainland China. Of the total 134 accounts, the number of Weibo accounts in the eastern region (n=68, 50.7%) is higher than those in the central region (n=30, 22.4%) and the western region (n=36, 26.9%). There are a total of 158 CDC facilities in the eastern region, and 68 of these have registered government Weibo accounts, with a registration rate of 43.0%. The highest registration rate is in the capital, Beijing (17/17, 100.0%), and the lowest is in Hainan (0/5, 0%). There are 112 CDC facilities in the central region, and 30 of them have registered Weibo accounts, with a registration rate of 26.8%. Henan Province (10/18, 55.6%) has the highest rate, and Heilongjiang Province (0/14, 0%) has the lowest rate. The number of CDC facilities in the western region is 180, of which 36 have registered Weibo accounts, with a registration rate of 20.0%. In the western region, Ningxia (3/6, 50.0%) has the highest registration rate, while Tibet (0/8, 0%) and Qinghai (0/9, 0%) have the lowest rate.

Table 1. Distribution of CDC facilities and the CDC's government Weibo accounts in mainland China.

Province	CDC[a], n	Government Weibo accounts, n (%)	Registration rate for each location (%)
Total	450	134 (100)	29.8
Eastern region	158	68 (50.7)	43.0
Beijing	17	17 (12.7)	100.0
Tianjin	17	7 (5.2)	41.2
Hebei	12	3 (2.2)	25.0
Liaoning	15	11 (8.2)	73.3
Shanghai	17	5 (3.7)	29.4
Jiangsu	14	6 (4.5)	42.9
Zhejiang	12	8 (6.0)	66.7
Shandong	17	6 (4.5)	35.3
Guangdong	22	4 (3.0)	18.2
Fujian	10	1 (0.7)	10.0
Hainan	5	0 (0)	0
Central region	112	30 (22.4)	26.8
Hubei	14	4 (3.0)	28.6
Hunan	15	5 (3.7)	33.3
Henan	18	10 (7.6)	55.6
Anhui	17	5 (3.7)	29.4
Jiangxi	12	1 (0.7)	8.3
Shanxi	12	4 (3.0)	33.3
Jilin	10	1 (0.7)	10.0
Heilongjiang	14	0 (0)	0
Western region	180	36 (26.9)	20.0
Guangxi	15	2 (1.5)	13.3
Chongqing	39	3 (2.2)	7.7
Sichuan	22	8 (6.0)	36.4
Guizhou	10	2 (1.5)	20.0
Inner Mongolia	13	5 (3.8)	38.5
Yunnan	17	3 (2.2)	17.6
Tibet	8	0 (0)	0
Shaanxi	11	2 (1.5)	18.2
Gansu	15	5 (3.8)	33.3
Qinghai	9	0 (0)	0
Ningxia	6	3 (2.2)	50.0
Xinjiang	15	3 (2.2)	20.0

[a] CDC: Center for Disease Control and Prevention.

Time Distribution of the CDC's Government Weibo Account Adoption

The CDC's government Weibo account adoption has a trend of increasing year by year, but there are still some provinces that have not registered government Weibo accounts. The first CDC facility to register a Weibo account in mainland China was the CDC in Lianyungang, Jiangsu Province, which registered on January 17, 2011. Since then, the registration rate of government Weibo accounts has increased year by year. In 2011, of the total 450 CDC facilities, 22 CDC facilities in 14 administrative regions registered government Weibo accounts, with a registration rate of 4.9% (Figure 3). From 2012 to 2014, there were 80 additional CDC facilities with registered government

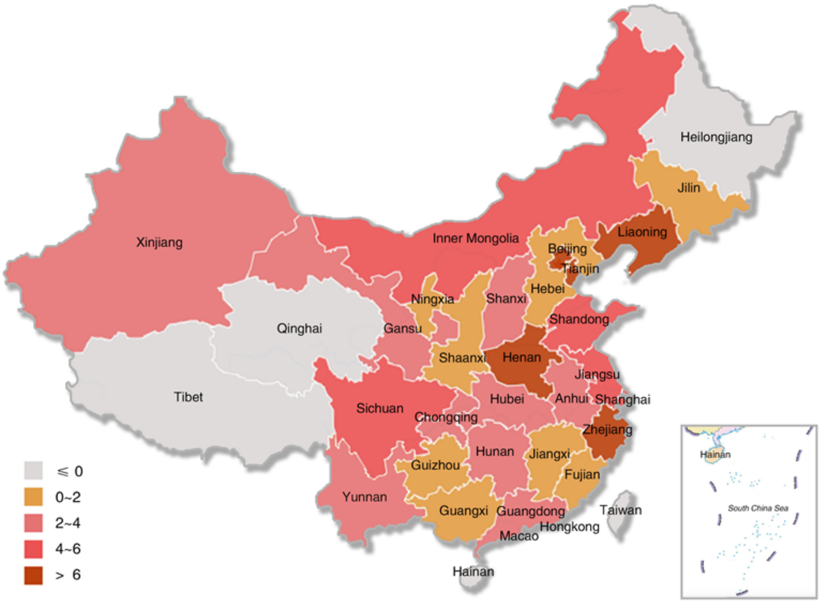
Weibo accounts, increasing the registration rate to 22.7% (102/450), and overall, the accounts were distributed in 26 provincial administrative regions (Figure 4). From 2015 to 2017, there were 24 additional CDC facilities with registered government Weibo accounts, with a registration rate of 28.0% (126/450), and overall, the accounts were distributed in 26 provincial administrative regions (Figure 5). From 2018 to March 2019, there were 8 new CDC facilities with registered government Weibo accounts, increasing the registration rate to 29.8% (134/450), and overall, the accounts were distributed in 29 provincial administrative regions (Figure 6). Currently, the 3 provincial administrative regions of Heilongjiang, Qinghai, and Hainan have not registered government Weibo accounts.

Figure 3. Diffusion of government Weibo accounts (2009-2011).

Figure 4. Diffusion of government Weibo accounts (2012-2014).

Figure 5. Diffusion of government Weibo accounts (2015-2017).

Figure 6. Diffusion of government Weibo accounts (2018-2019).



Hierarchical Distribution of the CDC's Government Weibo Adoption

The central-level CDC registered a government Weibo account later than the provincial-level CDC, and the provincial-level CDC registered later than the prefecture-level CDC, showing a "bottom-up" policy learning process. A total of 450 CDC facilities at the provincial and prefectural levels in China have registered Weibo accounts, of which the earliest one is the CDC in Lianyungang, Jiangsu Province, which is at the prefectural level. The registration time was January 17, 2011. On March 31, 2011, the first provincial CDC (Hunan CDC) registered a government Weibo account, which was later than the account registration time of the prefecture-level administrative district. The registration time of the official Weibo, "Science Popularization of Disease Control and Prevention," of the Chinese Centers for Disease Control and Prevention (as a central institution) was August 23, 2019, which was more than 9 years later than the prefecture-level CDC and provincial administrative regions where Weibo accounts were first registered. In addition, during this period, more than 100 CDC facilities at the provincial and prefectural levels in mainland China registered government Weibo accounts.

Operation of the CDC's Government Weibo Accounts

The CDC's Government Weibo Accounts Operating Certification

The Blue V certification is how Sina Weibo authenticates government, media, institutional, and other official accounts, as shown in the bottom right of the Weibo profile photo in Figure 1. This logo shows that Sina Weibo has verified that an account is an organization's official account, and the main body of the account is more authoritative and authentic, which can help the public to accurately identify official accounts. Of the 134 CDC facilities that have registered Sina Weibo accounts, 88.1% (n=118) of the accounts have been certified with Blue Vs, and the average number of Weibo followers with V-certified accounts is 18,753. There are 16 CDC accounts without certification, and followers do not pay attention to the non–V-certified accounts. One exception is an account that has 12,514 followers; the other Weibo accounts have less than 500 followers.

Dropout in the Use of the CDC's Government Weibo Accounts

The 134 CDC Weibo accounts have different degrees of use and dropout (Table 2). A few of them are "zombie microblogs," that is, 3.7% (n=5) of the accounts have not posted any content since registering their Weibo account. Some of the accounts are inactive. Only 37.3% (n=50) of the accounts had posted tweets in the past 30 days, 7.5% (n=10) of the accounts had posted tweets in the last 31-90 days, 16.4% (n=22) of the accounts had posted tweets in the last 91-365 days, and 35.1% (n=47) of the accounts had not posted tweets in more than 1 year. Among the latter, most accounts are in the eastern region, accounting for 15.7% (n=21), which is higher than those from the western region at 11.2% (n=15) and those from the central region at 8.2% (n=11). It can be seen that, although the registration rate in the eastern region is relatively high, the dropout rate of more than 1 year is also relatively high.

Table 2. The update time of the most recent Weibo tweets from the Center for Disease Control and Prevention's government accounts (n=134).

Last update time (days)	Eastern region, n (%)	Central region, n (%)	Western region, n (%)	Total, n (%)
≤30	27 (20.1)	12 (9.0)	11 (8.2)	50 (37.3)
31-90	6 (4.5)	1 (0.7)	3 (2.2)	10 (7.5)
91-365	13 (9.7)	5 (3.7)	4 (3.0)	22 (16.4)
>365	21 (15.7)	11 (8.2)	15 (11.2)	47 (35.1)
No content	1 (0.7)	1 (0.7)	3 (2.2)	5 (3.7)

Followers of the CDC's Government Weibo Accounts Were Polarized

The total number of followers on the CDC's government Weibo accounts was 3,588,544, the average was 26,780 (SD 165,506), and the median was 496. Among the accounts, the Changsha CDC had the largest number of followers, with a total of 1,357,440 followers. The one with the least number of followers was the Zhangye CDC, with a total of 2 followers. The total number of followers for 7 of the CDC's government Weibo accounts was less than ten; 17 of the CDC's government Weibo accounts had more than 10,000 followers; and 3 of the CDC's government Weibo accounts had more than 100,000 followers. These 3 accounts were the Hebei CDC (n=151,289), the Hunan CDC (n=1,331,173), and the Changsha CDC (n=1,357,440).

Interaction of the CDC's Government Weibo Accounts

Reply Rate to Comments on the CDC's Government Weibo Accounts

Among the 1215 tweets selected for content analysis in this study, only 12 of the public comments received replies from the government, accounting for less than 1.0%. Statistical analysis of the reply rate and the reply time of the 12 replies found that 50.0% (n=6) of the replies to the comments on Weibo were in the form of "reply only once" and that the remaining 50.0% (n=6) were in the form of "interactive reply." The response time of 66.7% (n=8) of the Weibo comments was between 0 hours and 1 hour, 8.3% (n=1) of the response times to Weibo comments were between 1 hour and 8 hours, and 25.0% (n=3) of the response times to Weibo comments were more than 12 hours.

Influence of the Topic Type on the Interaction

Among the 1215 tweets across all topics, popularizing health knowledge had the most, reaching 606 (49.9%) tweets (Table 3). Disease control information and popularization of disease knowledge both accounted for more than 10.0%. The number of tweets about policy interpretation, emergency response, and Weibo help and citizen consultation had the least, with the number of posts accounting for 0.6% (n=7), 1.2% (n=15), and 1.2% (n=14), respectively. In terms of the number of Weibo retweets, comments, and likes, emergency information posts ranked first, with each post being retweeted 4.1 times, commented on 2.9 times, and liked 4.0 times on average. Policy interpretation received the least number of comments, all of which were 0. This shows that communication effect is better here than for other topic types in dealing with public health emergencies, and it has become the platform for interaction between the government and the public.

Table 3. Descriptive statistical analysis results based on the topic type of tweets.

Topic type	Disease control information	Emergency information	Popularization of health knowledge	Popularization of disease knowledge	Radiation hygiene/school hygiene	Government affairs trends	Policy interpretation	Weibo help/citizen consultation	Other
Tweets (n=1215), n (%)	187 (15.4)	15 (1.2)	606 (49.9)	137 (11.3)	24 (2.0)	95 (7.8)	7 (0.6)	14 (1.2)	130 (10.7)
Number of retweets, average	0.4	4.1	0.6	0.7	1.3	0.2	0	1.2	0.3
Number of comments, average	1.0	2.9	0.4	0.3	0.3	0.4	0	1.6	0.3
Number of likes, average	0.7	4.0	0.3	0.5	0.9	0.6	0.9	0.4	0.6

Influence of the Content Form on the Interactive Effect

Among the 1215 tweets, the number of posts with "only original text" was 222 (18.3%). The average numbers of retweets, comments, and likes with "only original text" posts were 0.3, 0.3, and 0.3, respectively. On the other hand, there were 993 (81.7%) posts with "pictures/videos/hyperlinks." The average numbers of retweets, comments, and likes with "pictures/videos/hyperlinks" posts were 0.6, 0.6, and 0.5, respectively. It can be seen that the average numbers of retweets, comments, and likes of microblogs with "pictures/videos/hyperlinks" were higher than those of microblogs with original text, and the interactive effect was better.

Influence of the Original Post on the Interactive Effect

Among the 1215 tweets, the number of "original posts" was 703 (57.9%), and the average numbers of retweets, comments, and likes for "original posts" were 0.6, 0.6, and 0.7, respectively. Additionally, the average number of "retweet posts" was 512 (42.1%), and the average numbers of retweets, comments, and likes for "retweeted posts" were 0.5, 0.4, and 0.2, respectively. It can be seen that the average number of retweets, comments, and likes for original posts were higher than those of retweeted posts, indicating that the original content was more in line with the public's preference.

Performances of the CDC's Government Weibo Accounts During COVID-19

Activity Rate of the CDC's Government Weibo Accounts

Of the 134 Weibo accounts, 15.7% (n=21) had a high level of activity, posting more than 5 tweets per day, and 11.2% (n=15) were moderately active, with 1-2 tweets per day. In addition, 23.1% (n=31) posted 1-10 tweets per month. These three categories add up to exactly 50%. However, 46.3% (n=62) of the CDC's government Weibo accounts have not been updated for more than 1 year, an increase of 11.2% (n=15) over the number of accounts before the epidemic.

Main Content of the CDC's Government Weibo Accounts: COVID-19

Compared with the pre-epidemic statistics, there were approximately 60,000 new tweets and 1.4 million new followers on the 134 CDC accounts. Of the tweets, 90% were about public health emergencies related to COVID-19. These tweets can be divided into four categories. The first is updating the daily epidemic situation including new confirmed cases, new deaths, new suspected cases, new asymptomatic infected people, cumulative cured and discharged cases, life and medical tracking of confirmed cases, etc. The second was announcing the CDC's work such as the details of procurement announcements. The third was educating the public about the epidemic, including the issuance of protection guidelines for specific places such as schools, companies, shopping malls, and subways; protection guidelines for specific groups such as pregnant women, couriers, taxi drivers, and sanitation workers; nutritional dietary guidelines during the epidemic; and the popularization of disinfectants and protective products. The fourth was publicizing the typical deeds and dedication of the antiepidemic pioneers, especially the CDC.

Interaction of the CDC's Government Weibo Accounts: Decreased Month by Month

At the beginning of the pandemic, the number of posts, comments, and likes was the highest. For example, on January 24, 2020, the tweet "A Letter from Beijing CDC to friends from all over the country who come (return) to Beijing" was released by the Beijing CDC, and the total number of posts, comments, and likes was more than 2000. With the normalization of epidemic prevention and control, the amount of interaction is decreasing. During the epidemic, the questions that the public responded to in the comment area were broadly divided into the following three categories: epidemic prevention and control policies, specific information about new cases, and praising the CDC for its efforts. However, the CDC's government Weibo accounts still rarely respond to public comments, and only a few of the CDC's government Weibo accounts provide an office phone number, which is similar to the results of previous studies.

Discussion

Principal Findings

An increasing number of public health authorities in China have actively adopted new information platforms and tools for information disclosure and communication, which is the first step in improving communication between the government and the public.

First, the registration rates of the Chinese CDC's government Weibo accounts in the central and western regions are lower than that in the eastern region. This may be influenced by the government's motivation and ability to adopt new technologies. According to the "motivation-capability" framework, whether the government adopts new technology or not mainly depends on the motivation and ability of the government; only when that motivation and innovation ability are strong will the new technology be used [23]. Many studies have shown that there is a positive correlation between the level of economic development and the level of government information development [24,25]. Compared with the eastern region, the central and western regions of China are at a geographical disadvantage in terms of economic development, openness, and financial resources [22], so their ability to use government Weibo accounts is also relatively low.

Second, the diffusion process of social media adoption among China's public health authorities presents two characteristics: one is preferential diffusion among neighboring provinces, that is, horizontal learning and imitation, and the second is the vertical diffusion of local policies from the bottom to the top. Previous studies have shown that due to the influence of the "neighborhood effect," it is easier for a government to follow the examples of the neighboring leading regions' governments [26] and learn from the similar experiences of neighboring governments, which helps to improve the success rate of innovation and effectively avoid risks [27]. Previous studies have also confirmed the positive role of the policy learning process [27]. The central government also takes the policy innovation of the local government as the source of policy learning, and once the local policy is successful, it will revise the corresponding policy in time [28].

Third, the social media operations of Chinese public health authorities are still in a passive state. Although nearly 90% of the accounts have official authentication, which can help the public to quickly identify official accounts, and some accounts have a strong ability to reach followers, the overall activity of the accounts was low. Previous studies have shown that the more government is involved in social media operations, the higher the public's expectation of government interaction [4]. A negative operational status is likely to dampen the public's enthusiasm for online participation and may not even live up to the public's relationship expectations [29]. Only by continuously and actively operating social media can we better maintain a normal relationship between public health authorities and the public. Therefore, once the government registers a social media account, it must maintain their social media activity and update daily information frequently.

Fourth, the use of social media by Chinese public health authorities is more inclined to be one-way information dissemination such as popularizing health knowledge, while two-way communication with the public is still limited. For China, where scientific literacy is generally low, the popularization of basic health knowledge is important, but what is more important is how the government mobilizes and communicates using social media to encourage the public to participate in dialogue and cooperation. Especially in the case of limited traditional communication channels, the role of social media is more prominent. Previous studies have indicated that the new dimension that social media brings to the field of public health is that it can change the nature and speed of the interaction between the public and public health authorities [30]. Therefore, governments should use social media not only as a channel to release public health information and transmit health information to the public promptly but also to have two-way dialogues with the public to increase public participation in all stages. This will allow social media to become the best practice for improving communication between the government and the public.

Fifth, the use of social media by Chinese public health authorities provided an important channel of information disclosure and communication for the public during COVID-19, and it generally performed better than before the epidemic, although it still fell short of the Chinese government and public's requirements. The State Council of China requires that "public messages on the government Weibo should be carefully reviewed, released and processed" [31], but the CDC's government Weibo accounts tend to be "one-way" by informing the public of the latest developments of the epidemic, and they fail to respond to public inquiries and the large amount of misinformation during the epidemic in a timely manner. In addition, this study shows that the social media interaction effect during the period from January to June 2020 showed a declining trend. This is also consistent with previous research that the government and the public discussion trends in social media can predict the evolution of an epidemic's dynamics [15]. As the epidemic becomes normalized, the public interest in the dynamics of the epidemic, control policies, and guidelines for prevention and control is waning [32].

Moreover, whether we can maximize the function of social media for public health authorities also depends on the changes of the administrative system and political culture. In China's centralized political system, the government is dominant in the relationship between the government and the public, and China's current top-down decision-making and execution mechanism has many bureaucratic levels, which are not conducive to the effective transmission of information. In addition, the common people usually hold the idea of it being "difficult to deal with the government" and are reluctant to communicate with the government; therefore, there is a large psychological distance between the government and the public [20]. Social media provides an opportunity to improve the interaction between the government and the public. Chinese public health authorities must break the thinking mode of the "official standard," rethink the boundary between the government and the public, and promote the harmonious development of relations between the government and the public.

Limitations

There are some limitations in this study. First, the survey samples of this study did not include samples of county-level administrative regions (county-level administration regions are governed by prefecture-level administrative regions). Future research can study samples of the county-level administration regions and expand the research results. Second, this study only evaluates the government Weibo accounts; however, government Wechat accounts, as an emerging government social media platform, also have research value. Third, this study only uses descriptive statistics and content analysis, and did not investigate the psychology and behavior of the audience. Future research can use questionnaires, interviews, and other methods to further explore government Weibo use by the audience.

Conclusions

This study examines the current situation and interaction of social media use by public health authorities in China, a non-Western democratic country. This study analyzes the CDC's government Weibo accounts for the provincial- and prefecture-level administrative regions in mainland China, and explores how the public health authorities in China improve communication between the government and the public through social media. The results show that the adoption of government Weibo accounts has an uneven regional geographical distribution, steady diffusion year by year over time, and hierarchical bottom-up diffusion. Regarding the operations of government Weibo accounts, nearly 90.0% of government Weibo accounts have official certification, but there are dropouts in the specific operating process. One-third of the accounts have not provided updates for more than 1 year, and the number of microblog followers is polarized, with a maximum and minimum difference of 1 million. Regarding the interaction of government Weibo accounts, although the government Weibo accounts have changed the original layer-by-layer communication mode making communication between the government and the public more convenient, the Chinese government currently is more inclined to release one-way information, and the interaction with the public is limited. The response rate to comments was less than 1%. In terms of the influencing factors, emergency information, multimedia content, and original content are more helpful to promote communication between the government and the public. In the event of a public health emergency such as COVID-19, these accounts can function by updating epidemic information and protection information for the public, although there is still a gap in the two-way interaction. In general, government Weibo use is the first step in improving communication between the government and the public, but the effect is limited and needs to be improved.

Acknowledgments

This research was funded by the National Social Science Fund of China, grant number 18BZZ078.

Authors' Contributions

RZ conceptualized the study and reviewed and edited the manuscript. ML performed the data analysis. RZ and ML designed the methodology and prepared the original manuscript draft. Both authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Kott A, Limaye RJ. Delivering risk information in a dynamic information environment: framing and authoritative voice in Centers for Disease Control (CDC) and primetime broadcast news media communications during the 2014 Ebola outbreak. Soc Sci Med 2016 Nov;169:42-49 [FREE Full text] [doi: 10.1016/j.socscimed.2016.09.029] [Medline: 27682022]
2. Zeng R, Zhu D. A model and simulation of the emotional contagion of netizens in the process of rumor refutation. Sci Rep 2019 Oct 02;9(1):14164. [doi: 10.1038/s41598-019-50770-4] [Medline: 31578405]
3. Wang G, Zhang W, Zeng R. WeChat use intensity and social support: The moderating effect of motivators for WeChat use. Comput Hum Behav 2019 Feb;91:244-251. [doi: 10.1016/j.chb.2018.10.010]
4. Mergel I. A framework for interpreting social media interactions in the public sector. Government Inf Q 2013 Oct;30(4):327-334. [doi: 10.1016/j.giq.2013.05.015]
5. Cho SE, Park HW. Government organizations' innovative use of the internet: the case of the Twitter activity of South Korea's Ministry for Food, Agriculture, Forestry and Fisheries. Scientometrics 2011 Sep 28;90(1):9-23. [doi: 10.1007/s11192-011-0519-2]
6. Bonsón E, Torres L, Royo S, Flores F. Local e-government 2.0: social media and corporate transparency in municipalities. Government Inf Q 2012 Apr;29(2):123-132. [doi: 10.1016/j.giq.2011.10.001]
7. Zeng R, Chen Y, Li H. Participation in collective action in China: unfair experience, interests involved, and political efficacy. Soc Behav Pers 2018 Sep 07;46(9):1561-1572. [doi: 10.2224/sbp.7117]
8. Muniz-Rodriguez K, Ofori SK, Bayliss LC, Schwind JS, Diallo K, Liu M, et al. Social media use in emergency response to natural disasters: a systematic review with a public health perspective. Disaster Med Public Health Prep 2020 Feb;14(1):139-149. [doi: 10.1017/dmp.2020.3] [Medline: 32148219]
9. Ma L. What drives the adoption of social media applications by the public sector?: Evidence from local health departments. Int J Public Adm Digital Age 2016;3(4):76-93. [doi: 10.4018/ijpada.2016100106]
10. Harris JK, Mueller NL, Snider D. Social media adoption in local health departments nationwide. Am J Public Health 2013 Sep;103(9):1700-1707. [doi: 10.2105/AJPH.2012.301166] [Medline: 23865660]
11. Liu B, Kim S. How organizations framed the 2009 H1N1 pandemic via social and traditional media: implications for U.S. health communicators. Public Relations Rev 2011 Sep;37(3):233-244. [doi: 10.1016/j.pubrev.2011.03.005]
12. Sesagiri Raamkumar A, Tan SG, Wee HL. Measuring the outreach efforts of public health authorities and the public response on Facebook during the COVID-19 pandemic in early 2020: cross-country comparison. J Med Internet Res 2020 May 19;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]
13. Rahim AIA, Ibrahim MI, Salim FNA, Ariffin MAI. Health information engagement factors in Malaysia: a content analysis of Facebook use by the Ministry of Health in 2016 and 2017. Int J Environ Res Public Health 2019 Feb 18;16(4):591 [FREE Full text] [doi: 10.3390/ijerph16040591] [Medline: 30781691]
14. Vijaykumar S, Meurzec RW, Jayasundar K, Pagliari C, Fernandopulle Y. What's buzzing on your feed? Health authorities' use of Facebook to combat Zika in Singapore. J Am Med Inform Assoc 2017 Nov 01;24(6):1155-1159. [doi: 10.1093/jamia/ocx028] [Medline: 28449047]
15. Chen S, Xu Q, Buchenberger J, Bagavathi A, Fair G, Shaikh S, et al. Dynamics of health agency response and public engagement in public health emergency: a case study of CDC tweeting patterns during the 2016 Zika epidemic. JMIR Public Health Surveill 2018 Nov 22;4(4):e10827 [FREE Full text] [doi: 10.2196/10827] [Medline: 30467106]
16. Lee G, Kwak YH. An Open Government Maturity Model for social media-based public engagement. Government Inf Q 2012 Oct;29(4):492-503. [doi: 10.1016/j.giq.2012.06.001]
17. King D, Ramirez-Cano D, Greaves F, Vlaev I, Beales S, Darzi A. Twitter and the health reforms in the English National Health Service. Health Policy 2013 May;110(2-3):291-297. [doi: 10.1016/j.healthpol.2013.02.005] [Medline: 23489388]
18. Franco M, Tursunbayeva A, Pagliari C. Social media for e-government in the public health sector: protocol for a systematic review. JMIR Res Protoc 2016 Mar 11;5(1):e42 [FREE Full text] [doi: 10.2196/resprot.5421] [Medline: 26969199]
19. Cope JR, Frost M, Richun L, Xie R. Assessing knowledge and application of emergency risk communication principles among public health workers in China. Disaster Med Public Health Prep 2014 Jun;8(3):199-205 [FREE Full text] [doi: 10.1017/dmp.2014.29] [Medline: 24785302]

20. Jun K, Wang F, Wang D. E-Government use and perceived government transparency and service capacity. Public Performance Manage Rev 2014 Sep 27;38(1):125-151. [doi: 10.2753/pmr1530-9576380106]
21. CNNIC 45th Statistical Report on the Development of New Government Media. Sina Technology. 2020. URL: https://tech.sina.com.cn/i/2020-04-28/doc-iircuyvi0056994.shtml [accessed 2020-07-28]
22. Wei YD, Li H, Yue W. Urban land expansion and regional inequality in transitional China. Landscape Urban Plann 2017 Jul;163:17-31. [doi: 10.1016/j.landurbplan.2017.02.019]
23. Christensen CM, Anthony SD, Roth EA, Kaufman R. Seeing what's next: using the theories of innovation to predict industry change. Nonprofit Manage Leadership 2005 Apr;44(4):50-51. [doi: 10.1002/pfi.4140440412]
24. Ahn MJ. Adoption of e-communication applications in U.S. municipalities: the role of political environment, bureaucratic structure, and the nature of applications. Am Rev Public Adm 2010 Aug 18;41(4):428-452. [doi: 10.1177/0275074010377654]
25. Ahn MJ, Bretschneider S. Politics of e‐government: e‐government and the political control of bureaucracy. Public Adm Rev 2011;71(3):414-424. [doi: 10.1111/j.1540-6210.2011.02225.x]
26. Lee CP, Chang K, Berry FS. Testing the development and diffusion of e-government and e-democracy: a global perspective. Public Adm Rev 2011;71(3):444-454. [doi: 10.1111/j.1540-6210.2011.02228.x]
27. Volden C, Ting MM, Carpenter DP. A formal model of learning and policy diffusion. Am Polit Sci Rev 2008 Aug 01;102(3):319-332. [doi: 10.1017/s0003055408080271]
28. Mergel I. Social media adoption and resulting tactics in the U.S. federal government. Government Inf Q 2013 Apr;30(2):123-130. [doi: 10.1016/j.giq.2012.12.004]
29. Vanzetta M, Vellone E, Dal Molin A, Rocco G, De Marinis MG, Rosaria A. Communication with the public in the health-care system: a descriptive study of the use of social media in local health authorities and public hospitals in Italy. Ann Ist Super Sanita 2014;50(2):163-170. [doi: 10.4415/ANN_14_02_10] [Medline: 24968916]
30. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
31. Opinions of the General Office of the State Council of the People's Republic of China on promoting the healthy and orderly development of new government media. General Office of the State Council of the People's Republic of China. 2018. URL: http://www.gov.cn/zhengce/content/2018-12/27/content_5352666.htm [accessed 2020-07-28]
32. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: infodemiology study on social media data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]

Abbreviations

CDC: Center for Disease Control and Prevention

Edited by G Eysenbach; submitted 19.04.20; peer-reviewed by H Zhang, W Zhang; comments to author 11.07.20; revised version received 29.07.20; accepted 30.10.20; published 02.12.20.

Please cite as:
Zeng R, Li M
Social Media Use for Health Communication by the CDC in Mainland China: National Survey Study 2009-2020
J Med Internet Res 2020;22(12):e19470
URL: https://www.jmir.org/2020/12/e19470
doi:10.2196/19470
PMID:33151892

Original Paper

Using Social Media Data to Understand Consumers' Information Needs and Emotions Regarding Cancer: Ontology-Based Data Analysis Study

Jooyun Lee[1], PhD; Hyeoun-Ae Park[2], PhD, FAAN, FACMI; Seul Ki Park[2], PhD; Tae-Min Song[3], PhD

[1]College of Nursing, Gachon University, Incheon, Republic of Korea
[2]College of Nursing, Seoul National University, Seoul, Republic of Korea
[3]Department of Health Management, Sahmyook University, Seoul, Republic of Korea

Corresponding Author:
Hyeoun-Ae Park, PhD, FAAN, FACMI
College of Nursing
Seoul National University
103 Daehak-ro, Jongno-gu
Seoul, 03080
Republic of Korea
Phone: 82 2 740 8827
Fax: 82 2 766 1852
Email: hapark@snu.ac.kr

Abstract

Background: Analysis of posts on social media is effective in investigating health information needs for disease management and identifying people's emotional status related to disease. An ontology is needed for semantic analysis of social media data.

Objective: This study was performed to develop a cancer ontology with terminology containing consumer terms and to analyze social media data to identify health information needs and emotions related to cancer.

Methods: A cancer ontology was developed using social media data, collected with a crawler, from online communities and blogs between January 1, 2014 and June 30, 2017 in South Korea. The relative frequencies of posts containing ontology concepts were counted and compared by cancer type.

Results: The ontology had 9 superclasses, 213 class concepts, and 4061 synonyms. Ontology-driven natural language processing was performed on the text from 754,744 cancer-related posts. Colon, breast, stomach, cervical, lung, liver, pancreatic, and prostate cancer; brain tumors; and leukemia appeared most in these posts. At the superclass level, risk factor was the most frequent, followed by emotions, symptoms, treatments, and dealing with cancer.

Conclusions: Information needs and emotions differed according to cancer type. The observations of this study could be used to provide tailored information to consumers according to cancer type and care process. Attention should be paid to provision of cancer-related information to not only patients but also their families and the general public seeking information on cancer.

(J Med Internet Res 2020;22(12):e18767) doi:10.2196/18767

KEYWORDS

social media; ontology; cancer; health information needs; cancer information; emotion

Introduction

Dealing with cancer is both physically and mentally difficult, and patients require information on not only cancer itself but also on how to live with cancer [1-3]. There are time and situation constraints [3] that can hinder fulfillment of these requirements by health care providers such as physicians and nurses. Moreover, such information needs cannot be met by family members owing to their lack of expertise [2].

Health care consumers often use social media to exchange information, share experiences, and seek emotional support. They seek information from social media about diseases, treatments, and statistics to understand the disease and for help in making decisions. They also use social media to relieve anxiety and promote comfort by sharing their experiences and feelings, arising with cancer [1,4,5]. High-quality information is provided by cancer information portals, most of which are operated by government or professional societies. However, these portals are not designed to support sharing of experiences and feelings among patients. Therefore, consumers use social media to interact with each other by writing and reading about their shared experiences and feelings.

People use social media to share opinions, perceptions, concerns, and worries about health conditions [6]. Such posts have proven to be effective in identifying the interests and concerns of health care consumers related to the prevention, diagnosis, treatment, and management of diseases and the emotions related to diseases [7-10]. In recent years, there have been many studies [8,11,12] that extract health-related topics from social media data. The premise of these studies is that the topics posted on social media and their frequencies reflect the extent of consumers' health information needs [8]. A thorough understanding of these needs would be helpful in providing tailored information to consumers.

Text clustering [8] and machine learning [11,12] have been widely used to extract health-related topics from social media data. Lu et al [7] integrated medical terminology using the Unified Medical Language System to reflect the structure of medical knowledge in text clustering of messages posted by patients with lung cancer, breast cancer, and diabetes on online health communities. They were able to detect health-related topics effectively using this approach. The use of specific ontology for the domain of interest is helpful in the effective identification of relevant topics from social media data.

Although certain cancer-related ontologies, such as those for liver, breast, and gastric cancer, are available, they were developed with professional medical terms for data extraction from or integration with clinical databases [13-15] and are not suitable for analyzing social media data posted in consumer terms. Therefore, it is necessary to develop an ontology with terminology containing consumer terms to analyze social media data posted by consumers.

This study was conducted to develop a cancer ontology with terminology containing consumer terms and to analyze social media data to identify health information needs and emotions related to cancer.

Methods

The study consisted of 2 stages: (1) development and evaluation of a cancer ontology, and (2) analysis of social media data using the ontology.

Development and Evaluation of a Cancer Ontology

Ontology development was performed based on previous reports by Noy and McGuinness [16] and Jung et al [17].

First, the domain and scope of the ontology were determined using the following competency questions: (1) What types of cancer are mentioned in posts on social media? (2) Which care delivery processes (eg, prevention, diagnosis, treatment) of cancer in general and specific cancer types are mentioned in social media posts? (3) What cancer-related topics are mentioned in posts for cancer in general and specific cancer types? (4) What emotions, which mean a range of feelings a patient with cancer can experience when dealing with cancer, are mentioned in posts for cancer in general and specific cancer types?

The purpose of the ontology was determined as collecting and analyzing social media data to identify cancer information needs and emotions related to cancer.

Second, existing ontologies on cancer were identified: the Liver Cancer Ontology [13], Breast Cancer Ontology [14], and Gastric Cancer Ontology [15]. Each is limited to a specific type of cancer, and none includes consumer terms. Therefore, a new ontology was developed to include various types of cancer and consumer terms.

An existing ontology on emotion was also identified—the Sentiment Ontology for Social Web [18]. This ontology has top-level classes of emotion as positive, neutral, or negative. These top-level classes were too broad, and the second-level class was too detailed to describe a range of feelings a patient with cancer can experience when dealing with cancer. Therefore, a new ontology reflecting emotions accompanying cancer was deemed necessary.

Third, terms extracted from the 3 existing cancer-related ontologies, the Sentiment Ontology for Social Web, cancer information portals, and social media posts related to cancer were enumerated.

The cancer information portals that were reviewed to extract terms were 2 US websites (the National Cancer Institute [19] and American Cancer Society [20]), 1 UK website (Cancer Research UK [21]), and 2 Korean websites (the National Cancer Information Center [22] and National Health Information Portal by Korea Centers for Disease Control and Prevention [23]).

These portals included information on emotions accompanying cancer and how to manage them. Emotions included across portals were anger, guilt, and depression [19-22]. In addition, the Cancer Research UK website [21] included overwhelmed, denial, anxiety, fear, and sadness; the National Cancer Institute [19] included overwhelmed, denial, anxiety, fear, and sadness, loneliness, hope, and gratitude; and the Korean National Cancer Information Center [22] included overwhelmed, denial, anxiety, hope, and gratitude.

Natural language processing (NLP) was used to extract consumer terms from social media data on cancer. Terms with the same meaning as those extracted from existing cancer-related ontologies and cancer information portals or terms with a new meaning related to cancer were collected as consumer terms. These included (1) heteronyms such as "*jol-eob*" (*graduation* in English, meaning *complete cure*); (2) abbreviations such as "*jaegeom*" (an abbreviation of *jaegeomsa*, meaning *re-test* in English), "*chompa*" (an abbreviation of *cho-eumpa*, meaning *ultrasound* in English), "*holmon*" (an abbreviation of *holeumon*, meaning *hormone* in English); and (3) terms used for herbal medicine and complementary therapies.

Fourth, the classes and their hierarchy and relationships were defined. The collected terms were grouped according to semantic meaning and determined concepts as classes with an independent existence. Hierarchies of the classes were designed based on the relationships of the concepts. The superclass and subclass concepts of the ontology were determined by analyzing the structures of the ontology. A list of synonyms was compiled for each class concept as a terminology presenting the relationship between the concept and synonyms in the ontology.

Nine domains of cancer-related emotions were identified: *overwhelmed, denial, anger, fear and anxiety, sadness and depression, guilt, loneliness, hope,* and *gratitude*. Each domain was defined as a class concept, and a list of synonyms of the class concepts was mapped as a terminology.

Fifth, the structure, correctness, and quality of the ontology were evaluated using the evaluation tool described below and by interviewing 3 domain experts: 2 professors of family medicine and 1 professor of bioinformatics.

The tool consisted of 13 items selected from the studies of Hlomani and Stacey [24] and Kehagias et al [25]. The items were scored on a 5-point scale for structure (size, depth of hierarchy, breadth of hierarchy, balance, overall complexity, and connectivity between concepts), correctness (accuracy, completeness, conciseness, and consistency), and quality (computational efficiency, adaptability, and clarity). Interviews included open-ended questions that allowed experts to recommend revisions of the ontology. The ontology was revised based on the results of the evaluation.

Analysis of Social Media Data

Data Collection and Preparation

The social media data for this study were posts on cancer collected using a crawler from online communities and blogs of 4 social media platforms in South Korea, namely *Naver*, *Daum*, *Tistory*, and *Egloos*, between January 1, 2014 and June 30, 2017.

A total of 302 concepts and synonyms of the *cancer type* superclass were used as keywords for post extraction, and 418 concepts were used as stop keywords. For example, when certain Korean words or morphemes, such as "*agseong*" (meaning *malignant* in English) or "*am*" (meaning *cancer* in English) are combined with other words, the phrase could become a word or morpheme with a completely different meaning, such as "*agseong virus*" (meaning computer virus) or "*an-am*" (name of a district in Seoul). Posts containing 59 advertising keywords (eg, detoxification, antioxidant therapy, and enzyme therapy), suggesting an advertising post were removed.

A total of 754,744 posts were extracted from online communities and blog sites. When categorized by source into blogs and online communities, 442,669 (58.7%) were blog posts. Of the 754,744 posts, 234,118 were from 2014; 235,509 were from 2015; 200,553 were from 2016; and 84,564 were from the first half of 2017. Most of the posts (737,575; 97.7%) were from *Naver* and *Daum*, the 2 major social media platforms in South Korea.

Next, ontology-based NLP was performed on the posts to extract class concepts.

The data collection and NLP were carried out in collaboration with a Korean telecommunications company (Smart Insight). During NLP, identifying information (such as name, phone number, and account) was removed, and masked data were delivered to the research team.

Frequency Analysis of Posts

The unit of analysis was the post, and the frequencies of posts containing single specific class concepts were counted.

First, the relative frequencies of posts containing specific cancer types were counted and compared with the national cancer statistics of Korea. Top-ranked cancer types in social media posts were selected for further analyses. Second, the relative frequencies of posts containing superclass concepts were counted and compared by cancer type. Finally, the frequencies of posts containing end-node class concepts were counted and organized by cancer type.

None of the posts used in this study had any identifying information. The study was approved by the Institutional Review Board of Seoul National University (No. 1802/001-006).

Results

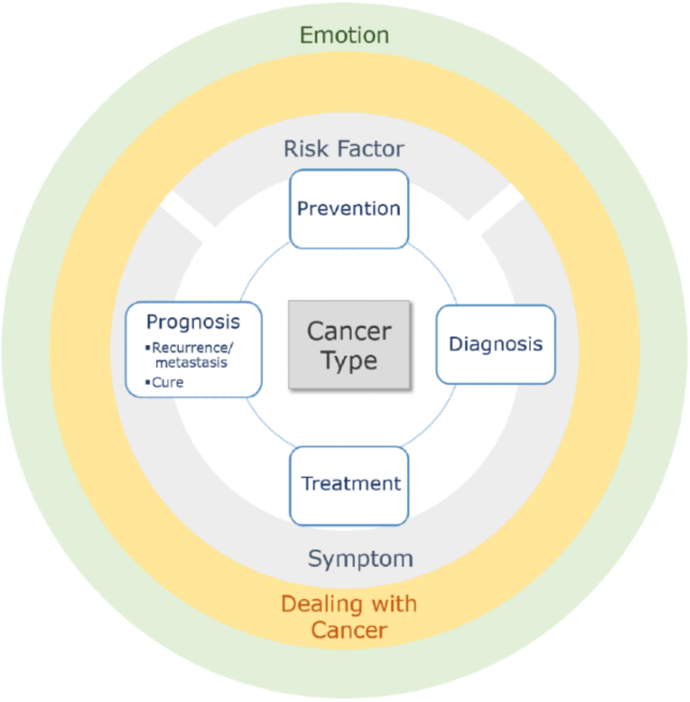
Development and Evaluation of the Ontology

Based on the existing cancer-related ontologies and cancer information portals, 10 superclasses were identified: *cancer type, prevention, diagnosis, treatment, prognosis* (including *recurrence* and *cure*), *risk factor, symptom, side effect, dealing with cancer,* and *emotion*. The ontology consisted of concepts that represented care delivery processes and patient outcomes, such as *prevention, diagnosis, treatment,* and *prognosis*, and another set of concepts that represented how consumers managed, felt, perceived, and acted in their personal lives, such as *risk factor, symptom, side effect, dealing with cancer,* and *emotion*.

The average scores by the 3 experts of 13 items designed to evaluate the structure, correctness, and quality of the ontology ranged from 4.33 to 5 on the 5-point scale. The score for correctness was the highest (mean 4.83), followed by structure (mean 4.67) and quality (mean 4.67). A suggestion made by one of the experts was to combine the superclass of *side effect* with that of *symptom*, moving cancer symptoms and treatment side effects into subclasses of the *symptom* superclass, because it is difficult to distinguish side effects from symptoms without context. Another suggestion was to add the national cancer support system as a subclass of the *dealing with cancer* superclass. The ontology was revised to reflect the comments made by the experts.

The revised ontology had 9 superclasses, 213 class concepts, and 4061 synonyms. It had 3 to 4 levels of hierarchy, with 36 first-level subclasses and 41 second-level subclasses (Figure 1).

Figure 1. Cancer-related ontology superclasses.



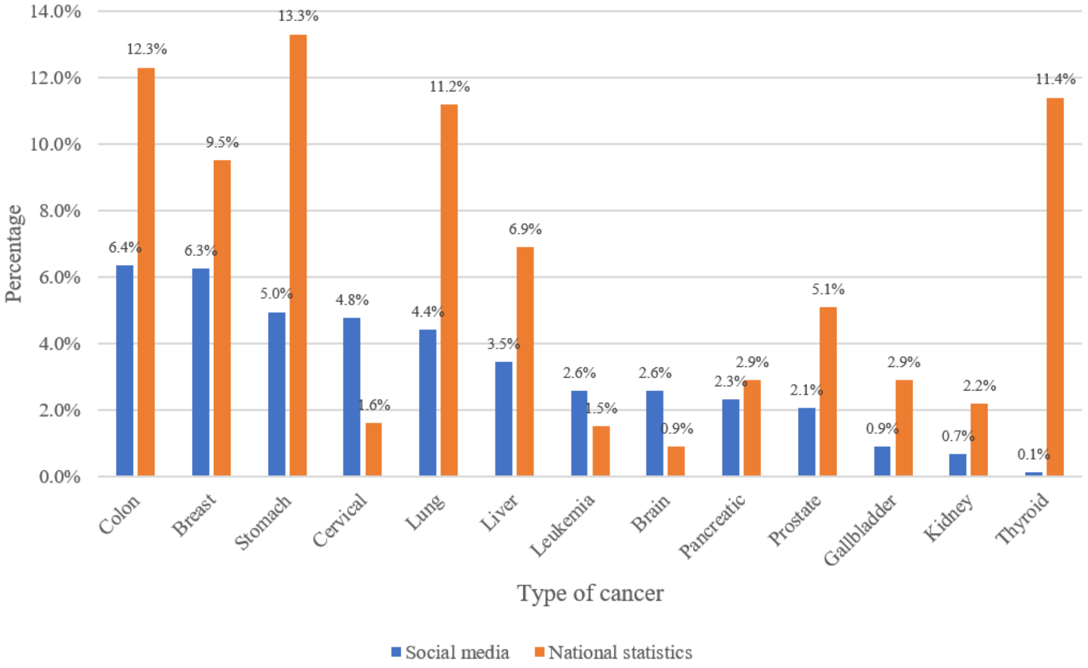
Analysis of Social Media Data

Comparison of Top-Ranked Cancer Types

Figure 2 presents the top-ranked cancer types mentioned on social media compared with national cancer statistics. Colon cancer (47,940/754,744, 6.4%) was the most frequently mentioned on social media, followed by breast cancer (47,235/754,744, 6.3%) and stomach cancer (37,378/754,744, 5.0%). The 4 highest-ranked cancers according to the national cancer statistics were stomach, colon, thyroid, and lung cancer.

Colon, breast, stomach, lung, liver, prostate, and pancreatic cancer were within the top 10 rankings in both social media data and the national cancer statistics. Cervical cancer, leukemia, and brain tumors, which ranked within the top 10 cancer types in the social media data, were not included in the top 10 cancer types in the national cancer statistics. Thyroid cancer ranked within top 3 in the national cancer statistics but was not included in the top-ranked cancer types on social media.

Figure 2. Comparison of top-ranked cancer types in social media posts and national cancer statistics.



Frequency of Posts Containing Superclass-Level Concepts

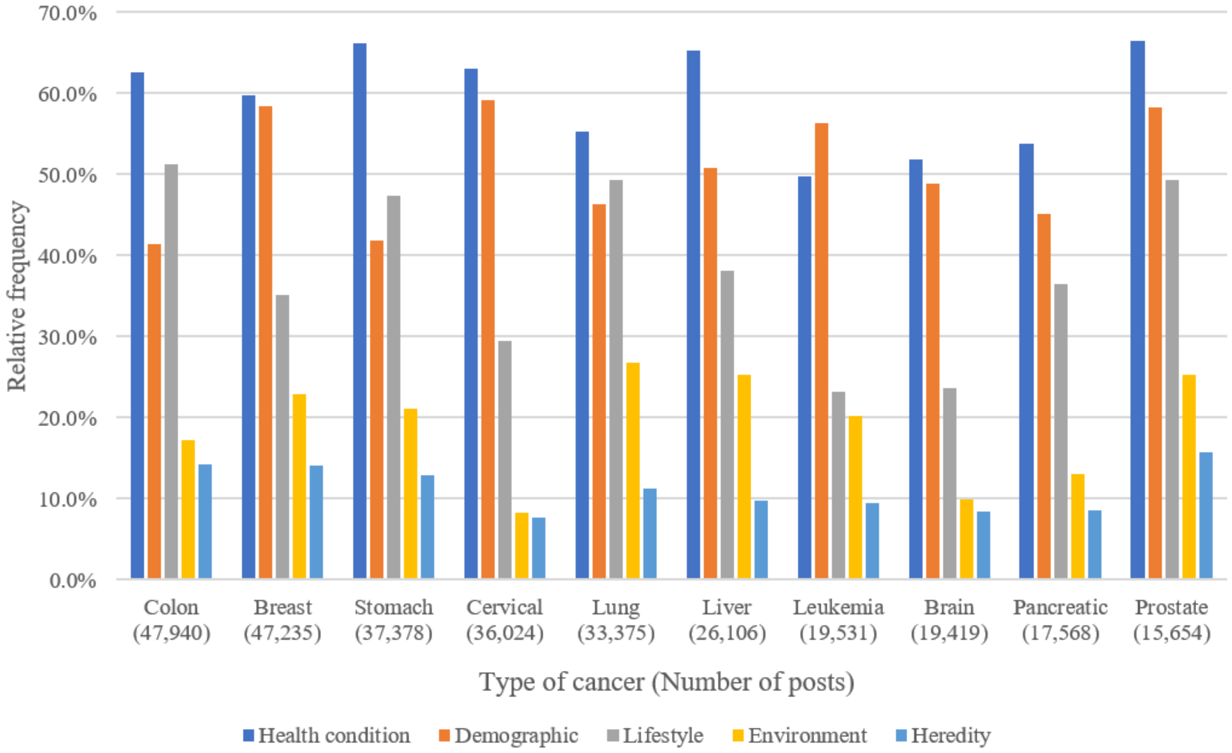
In terms of the frequency of posts at the superclass level, *risk factor* was the most frequent, appearing in 42.5% (320,568/754,744) of the posts, followed by *emotion* in 33.8% (254,920/754,744), *symptom* in 32.2% (243,010/754,744), *treatment* in 30.2% (227,942/754,744), *dealing with cancer* in 29.4% (221,996/754,744), *diagnosis* in 23.7% (178,498/754,744), *prevention* in 14.4% (108,408/754,744), and *prognosis* in 9.4% of the posts (70,583/754,744).

Risk Factor Superclass

Figure 3 shows the relative frequencies of posts containing *risk factor* superclass concepts for 10 specific cancer types. *Health condition-related risk factor* class concepts (such as related disease and health status) were the most common in most cancer types, followed by *demographic*, *lifestyle* (such as diet and exercise), *environmental*, and *hereditary risk factor* concepts.

Demographic risk factor class concepts appeared more frequently in posts on breast cancer, cervical cancer, leukemia, and prostate cancer than in those on other types of cancer. *Lifestyle-related risk factor* class concepts were more frequent in posts on colon, stomach, and lung cancer than in those on other cancer types. *Environmental risk factor* class concepts were more frequent in posts mentioning lung, liver, and prostate cancer than in those on other cancer types.

Figure 3. Relative frequencies of the risk factor superclass by top 10 cancer types in social media posts.

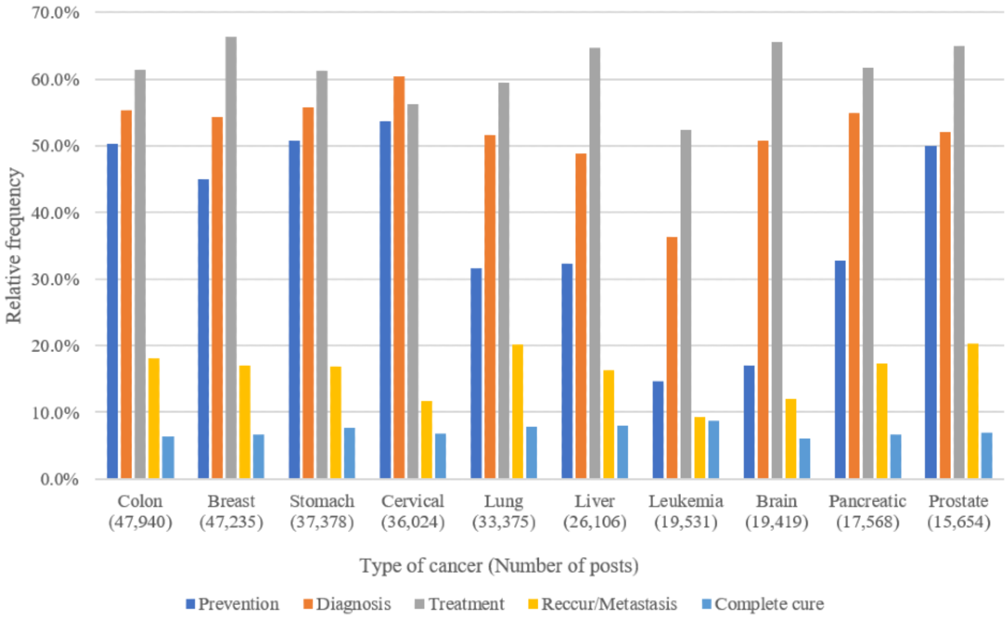


Care Process of Cancer: Prevention, Diagnosis, Treatment, Recurrence, and Cure Superclasses

Figure 4 shows the relative frequencies of posts containing superclass concepts related to the cancer care process by 10 specific cancer types. Posts containing *treatment* class concepts were the most common, followed by those related to *diagnosis*, *prevention*, *recurrence*, and *cure* class concepts for most cancer types, except for cervical cancer.

Prevention class concepts were more frequent in posts on cervical, stomach, colon, and prostate cancer and less frequent in posts on brain tumors and leukemia. The *diagnosis* class was dominant in cervical cancer posts, appearing more frequently than the *treatment* class. The *recurrence* class was the least frequently mentioned in leukemia posts, in which the *cure* class was common.

Figure 4. Relative frequencies of superclass concepts related to the care process of cancer by top 10 cancer types in social media posts.

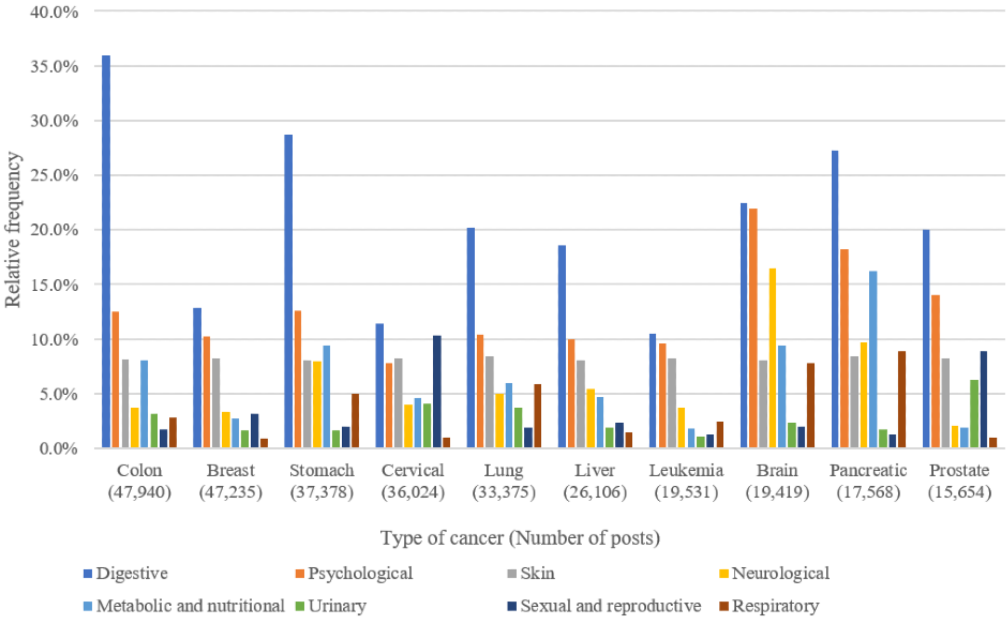


Symptom Superclass

Figure 5 shows the relative frequencies of posts containing *symptom* superclass concepts by 10 specific cancer types. *Digestive symptom* class concepts were the most common in posts on all cancer types, followed by *psychological symptom* concepts.

Digestive symptom class concepts were predominant in posts related to cancers of the digestive system, such as colon, stomach, and pancreatic cancer. *Psychological and neurological symptom* class concepts appeared more frequently in posts on brain tumors than in those on other cancers. *Metabolic symptom* class concepts were frequent in posts on pancreatic cancer, and *sexual and reproductive symptom* class concepts were frequent in posts on cervical and prostate cancer.

Figure 5. Relative frequencies of the symptom superclass by top 10 cancer types in social media posts.

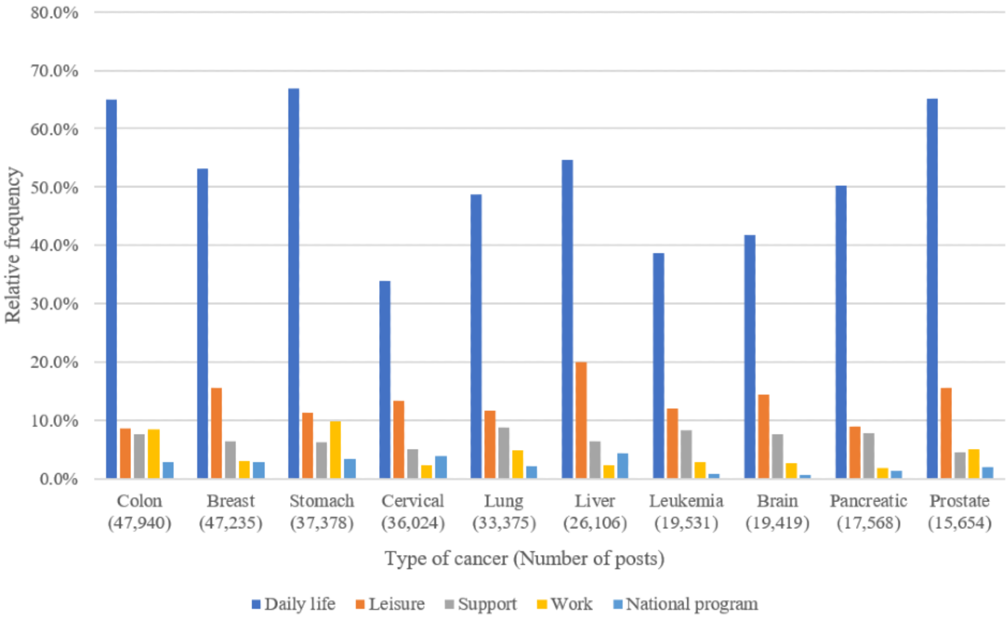


Dealing With Cancer Superclass

Figure 6 shows the relative frequencies of posts containing *dealing with cancer* superclass concepts by 10 specific cancer types.

Daily life class (involving diet and exercise) concepts were predominant in posts on the 10 cancer types. The most noticeable class concept in posts on *daily life* was *diet*. *Leisure* class concepts (involving sex life, travel, and driving) were more common in posts on liver and breast cancer.

Figure 6. Relative frequencies of the dealing with cancer superclass by top 10 cancer types in social media posts.

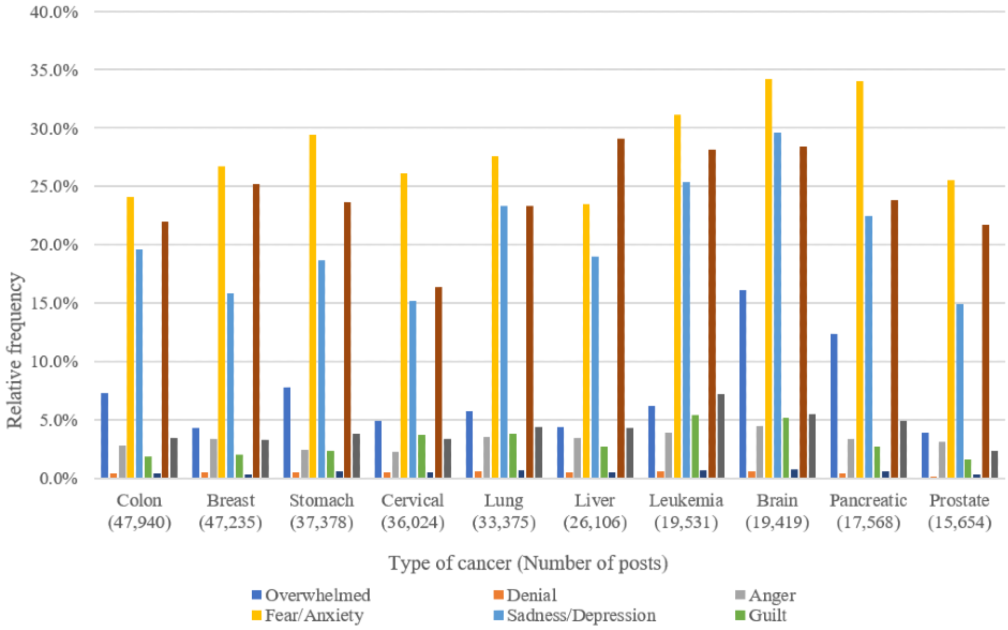


Emotion Superclass

Figure 7 shows the relative frequencies of posts containing *emotion* superclass concepts by 10 specific cancer types. *Fear/anxiety* class concepts were the most common, followed by *hope* and *sadness/depression* concepts, in posts on all cancer types except for liver cancer and brain tumors.

Hope class concepts were more frequently mentioned than *fear/anxiety* class concepts in posts on liver cancer. *Overwhelmed* class concepts were more frequent in posts on brain tumors and pancreatic cancer than in those on other cancers. *Gratitude* and *guilt* class concepts were more frequently mentioned in posts on leukemia than in those on other cancers.

Figure 7. Relative frequencies of the emotion superclass by top 10 cancer types in social media posts.

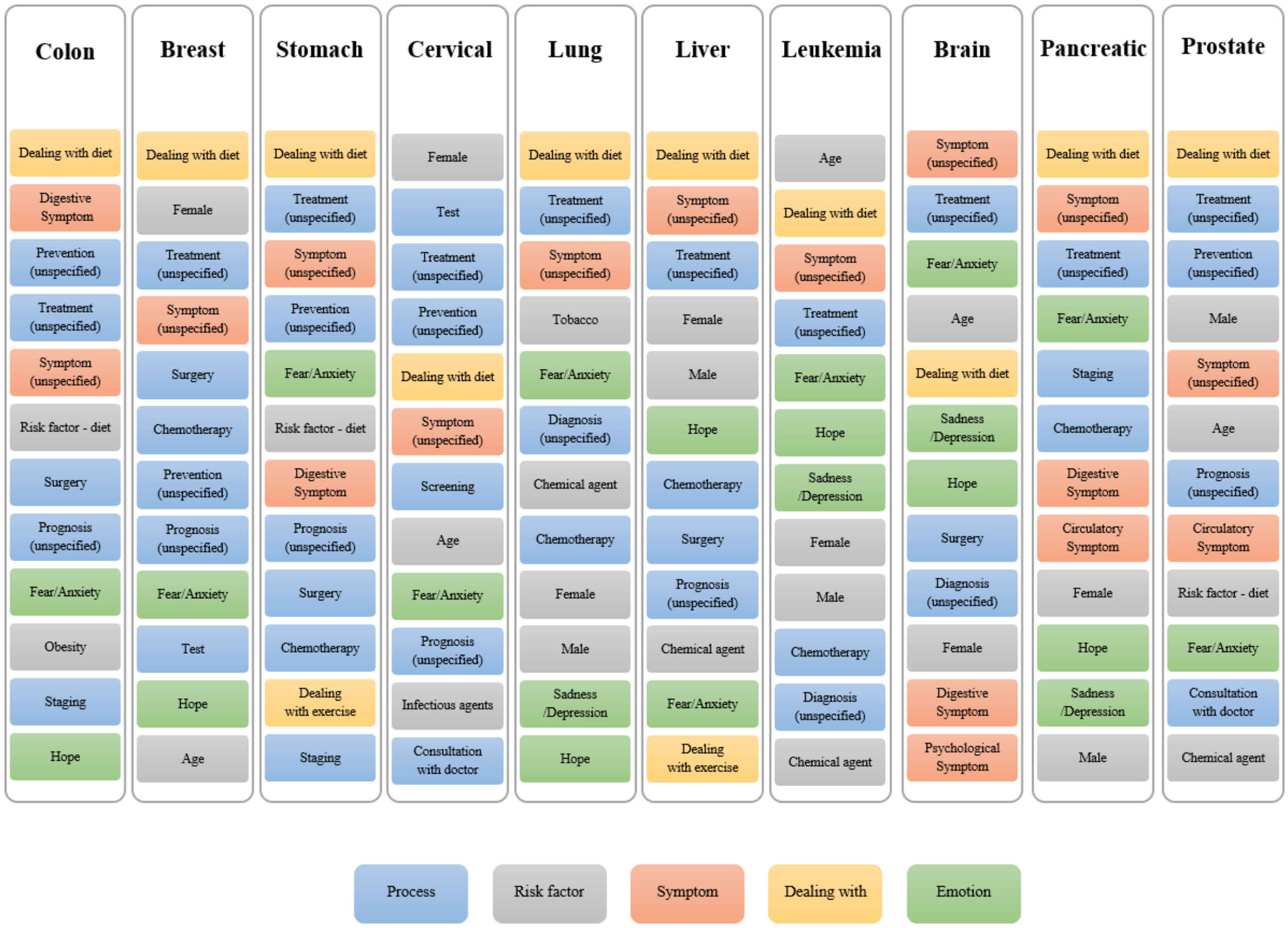


Frequencies of Posts Containing Class-Level Concepts by Cancer Type

Figure 8 shows the top-ranked end-node class concepts in order of relative frequency by cancer type.

The *dealing with diet* class was ranked first in posts related to 8 cancer types. The dominant concepts and terms appearing in posts about *dealing with diet* were food, protein, fruit, vitamin, meal, vegetable, and nutrition.

Figure 8. Top ranked end-node class concepts by cancer type.



Colon	Breast	Stomach	Cervical	Lung	Liver	Leukemia	Brain	Pancreatic	Prostate
Dealing with diet	Dealing with diet	Dealing with diet	Female	Dealing with diet	Dealing with diet	Age	Symptom (unspecified)	Dealing with diet	Dealing with diet
Digestive Symptom	Female	Treatment (unspecified)	Test	Treatment (unspecified)	Symptom (unspecified)	Dealing with diet	Treatment (unspecified)	Symptom (unspecified)	Treatment (unspecified)
Prevention (unspecified)	Treatment (unspecified)	Symptom (unspecified)	Treatment (unspecified)	Symptom (unspecified)	Treatment (unspecified)	Symptom (unspecified)	Fear/Anxiety	Treatment (unspecified)	Prevention (unspecified)
Treatment (unspecified)	Symptom (unspecified)	Prevention (unspecified)	Prevention (unspecified)	Tobacco	Female	Treatment (unspecified)	Age	Fear/Anxiety	Male
Symptom (unspecified)	Surgery	Fear/Anxiety	Dealing with diet	Fear/Anxiety	Male	Fear/Anxiety	Dealing with diet	Staging	Symptom (unspecified)
Risk factor - diet	Chemotherapy	Risk factor - diet	Symptom (unspecified)	Diagnosis (unspecified)	Hope	Hope	Sadness /Depression	Chemotherapy	Age
Surgery	Prevention (unspecified)	Digestive Symptom	Screening	Chemical agent	Chemotherapy	Sadness /Depression	Hope	Digestive Symptom	Prognosis (unspecified)
Prognosis (unspecified)	Prognosis (unspecified)	Prognosis (unspecified)	Age	Chemotherapy	Surgery	Female	Surgery	Circulatory Symptom	Circulatory Symptom
Fear/Anxiety	Fear/Anxiety	Surgery	Fear/Anxiety	Female	Prognosis (unspecified)	Male	Diagnosis (unspecified)	Female	Risk factor - diet
Obesity	Test	Chemotherapy	Prognosis (unspecified)	Male	Chemical agent	Chemotherapy	Female	Hope	Fear/Anxiety
Staging	Hope	Dealing with exercise	Infectious agents	Sadness /Depression	Fear/Anxiety	Diagnosis (unspecified)	Digestive Symptom	Sadness /Depression	Consultation with doctor
Hope	Age	Staging	Consultation with doctor	Hope	Dealing with exercise	Chemical agent	Psychological Symptom	Male	Chemical agent

Class concepts related to the care process of cancer were frequently highly ranked; the *unspecified treatment* class was highly ranked for all types of cancer. The *surgery* class was ranked within the top 10 class concepts in posts on colon, breast, stomach, and liver cancer, and brain tumors. The *chemotherapy* class was ranked within the top 10 class concepts in posts on breast, stomach, lung, liver, and pancreatic cancer, and leukemia.

Several risk factor–related class concepts ranked in the top 10 class concepts, but the order differed by cancer type. *Gender* and *age* classes frequently appeared in posts on breast, cervical, prostate cancer, and leukemia. The *dietary risk factor* class frequently appeared in posts on colon, stomach, and prostate cancer, and *obesity* class frequently appeared in posts on colon cancer. The *chemical risk factor* class frequently appeared in posts on lung, liver, and prostate cancer, and leukemia. The *infectious agent* class (such as viruses) frequently appeared in posts on cervical cancer.

There were no posts containing *support group/support community* class concepts in posts on thyroid cancer, but these concepts appeared in posts on all of the other top 10 cancers.

Discussion

Principal Findings

In this study, an ontology containing consumer terms was developed to collect and analyze social media data to identify health information needs and emotions related to cancer. The ontology has the following characteristics.

First, compared to other cancer ontologies, this ontology covers more cancer types and comprehensive topics, including the care process describing the interactions of health care providers and patients, risk factors that consumers have, emotions that consumers feel related to cancer, symptoms that consumers perceive, and the lifestyles that consumers lead.

Second, the ontology has a terminology component that presents consumer terms to analyze social media posts about cancer, including synonyms, heteronyms, and abbreviated expressions.

A total of 754,744 cancer-related posts on social media were collected. The 6-month collection of 2017 posts was less than half that of the full-year collection of the previous year. Monthly analysis showed cancer-related social media posts were more frequent in the second half of the year than in the first half of the year. In particular, the number of posts in February 2014 was almost twice the monthly average, which was the same as the search results of Google Trends. According to Google Trends, the cancer diagnosis and death from cancer of 2 Korean celebrities had been announced at this time.

Social media data were analyzed using NLP with ontology concepts and terms to identify consumers' information needs. The frequencies of topics in social media data may indirectly reflect consumers' information needs, as they frequently post on topics in which they are interested or about which they have concerns [8].

Thyroid cancer, which ranked in the top 3 in the national cancer statistics, was not included in the top-ranked cancer types on social media. This finding is similar to that of Buis and Whitten [26], showing that the information needs related to cancers with a low survival rate are higher than those for cancers with a high survival rate. The 5-year survival rate for thyroid cancer in Korea was about 100% in 2013-2017, compared to 70.4% for all types of cancer [22].

Frequency analysis by superclass revealed the highest frequency of posts on social media related to *risk factor* (320,568/754,744, 42.47%), followed by *emotion* (254,920/754,744, 33.78%), *symptom* (243,010/754,744, 32.20%), and *treatment* (227,942/754,744, 30.20%). These findings can be compared to those of 2 previous studies: Lu et al [7] clustered posts of patients with lung and breast cancer and extracted symptoms, examinations, and treatments (drugs, procedures) as dominant topics. Cho et al [1] performed a qualitative content analysis of Q & A posts of patients with breast cancer, and extracted treatment, physical condition, and lifestyle/self-care as dominant topics. They also found that 75% of the information requests included expressions of emotion [1]. Although these previous studies [1,7] focused mainly on the stages after diagnosis, such as treatment, examinations, and physical condition, our results revealed that the majority of consumers' information needs involved risk factors. Also, previous studies [1,7] used data collected from cancer community posts by patients with specific cancer types, whereas this study used data collected from social media posts by the general public. Social media data may include concerns of the public about cancer risk factors. In addition, as the ontology developed in this study covered comprehensive topics, including risk factors and emotions, this study had a more diverse focus than those of previous studies [1,7].

Consistent with the findings of the 2 previous studies [1,9], the most common emotions related to the top 10 cancers in this study were *fear/anxiety, hope*, and *sadness/depression*. Cho et al [1] reported that anxiety/worry, gratitude, fear, and sadness were frequent in posts by Korean women with breast cancer, and Freedman et al [9] reported that fears, anxiety, denial, and depression were frequent emotions cited on treatment in posts by patients with breast cancer. However, among positive emotions, gratitude was included in the top 3 in the study by Cho et al [1], whereas hope was ranked in the top 3 in our study. The question and answer board in Cho et al's study [1] included more posts on gratitude toward health care providers, in contrast to social media posts in this study that often included posts on hope for a cure by patients and their families.

Frequency analysis by the end-node class level indicated that *dealing with diet* ranked among the top-class concepts for most types of cancer. This finding was consistent with those of previous information-need studies [1,27] in Korean patients with breast cancer. Cho et al [1] reported that patients asked many questions related to diet, and Kim and Hur [27] reported high information-need scores for diet. These findings reflect the importance of diet in disease management that is perceived by the Korean population. Concepts and terms such as food, protein, fruit, vitamin, meal, vegetable, and nutrition were top ranked in the posts collected in this study, demonstrating consumers' information needs related to a healthy diet.

This study also compared the frequencies of superclass concepts and end-node class concepts by 10 types of cancer.

Regarding the care process of cancer, *treatment* was the most frequently mentioned for all types of cancer, except cervical cancer. *Diagnosis* and *prevention* appeared more frequently in posts on cervical cancer. In Korea, the national cancer screening program for cervical cancer is recommended from 20 years of age. Whereas breast and gastric cancer screenings are recommended from 40 years of age, and colorectal cancer screening is recommended from 50 years of age. As the opportunity to become interested in screening for cervical cancer comes earlier than for other cancers, the active use of social media by young women may have resulted in greater numbers of posts related to diagnosis and prevention.

The frequencies of risk factor–related class concepts, especially *diet, chemical, tobacco, obesity,* and *infectious agent*, differed between cancer types. Previous research findings have indicated that differences in post frequency on cancer topics reflect the different information needs for each cancer type [7,8]. Therefore, tailored information on risk factors should be provided according to consumers' needs by cancer type. In addition, tailored information on cancer can be provided according to the consumers' specific risk factors.

Posts on leukemia contained a higher rate of feelings of *guilt*. According to the National Cancer Information Center [22] data, leukemia has a higher incidence in children than those of other types of cancer. In many cases, posts related to leukemia were likely written by the parents of patients rather than by the patients themselves. This finding suggests that emotional management is necessary for not only patients with cancer but also their family members and friends.

Based on these findings, we suggest what information should be provided and how it can be provided. These suggestions would aid information providers, namely clinicians and portals operated by government or professional societies, to improve care for patients with cancer by providing relevant information based on consumers' information needs.

First, it is necessary to ensure that sufficient information on risk factors is provided to the public. Not only do information needs increase after cancer is diagnosed, but they are also high for risk factor management. Thus, information on risk factors should be provided depending on the concerns of consumers.

Second, the high consumers' information needs on healthy diets were noteworthy. Qualified information on diet should be provided to patients with cancer. Collaboration with a nutritionist would be effective in providing tailored nutritional information for each cancer type according to the needs of patients.

Third, in general, most information portals provide information through the same organization and the same flow of information for all types of cancer. Tailored information can be provided according to cancer type and the characteristics of the consumer, such as age, gender, and risk exposure. It is possible to make it easier to access the information that consumers want using keyword visualization or navigation. Different navigation routes could be applied according to cancer type. In addition, applying different types of visualization could improve the convenience of consumers before or after the cancer diagnosis.

Finally, providing information and emotional support are not separate but, instead, coexist. It is necessary to provide reliable information and management for emotional care so that people do not rely on only family and caregivers for emotional support. One possible approach is to combine the functions of online support groups with information portals to provide emotional support.

Limitations

This study had some limitations. Only the top 10 cancer types were analyzed. Future studies should analyze data on other types of cancer. The social media posts were made by not only patients with cancer but also caregivers and the general public. However, these populations could not be distinguished because identifying information of the consumers was not collected. Further research is needed to collect social media data with an identification algorithm to distinguish the status of the authors of the posts and to provide tailored information.

Conclusion

This study was performed to develop a cancer ontology with terminology containing consumer terms to collect and analyze social media data. The ontology consisted of 9 superclasses (*cancer type*, *prevention*, *diagnosis*, *treatment*, *prognosis*, *risk factor*, *symptom*, *dealing with cancer*, and *emotion*), 213 classes, and 4061 synonyms with consumer-generated terms. It used 9 emotional classes (*overwhelmed*, *denial*, *anger*, *fear and anxiety*, *sadness and depression*, *guilt*, *loneliness*, *hope*, and *gratitude*) to investigate emotional status in the social media data on cancer.

This ontology, containing comprehensive cancer-related topics, enabled identification and comparison of consumer interests and concerns about risk factors, dealing with cancer, and emotions as well as the care process in social media data. The results of this study showed that information needs and emotions differ according to cancer type. These observations could be used to provide tailored information to consumers according to the cancer type and care process. Care for patients with cancer can be improved by providing relevant information based on consumers' information needs.

Acknowledgments

This study was supported by grants from the National Research & Development Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (No.1720250) and the Gachon University research fund of 2018 (GCU-2018-0292).

Conflicts of Interest

None declared.

References

1. Cho J, Smith KC, Roter D, Guallar E, Noh D, Ford DE. Needs of women with breast cancer as communicated to physicians on the Internet. Support Care Cancer 2011 Jan;19(1):113-121. [doi: 10.1007/s00520-009-0794-0] [Medline: 21290791]
2. Hasson-Ohayon I, Tuval-Mashiach R, Goldzweig G, Levi R, Pizem N, Kaufman B. The need for friendships and information: Dimensions of social support and posttraumatic growth among women with breast cancer. Palliat Support Care 2016 Aug;14(4):387-392. [doi: 10.1017/S1478951515001042] [Medline: 26462622]
3. Mazor KM, Beard RL, Alexander GL, Arora NK, Firneno C, Gaglio B, et al. Patients' and family members' views on patient-centered communication during cancer care. Psychooncology 2013 Nov;22(11):2487-2495 [FREE Full text] [doi: 10.1002/pon.3317] [Medline: 23780672]
4. Corter AL, Speller B, Sequeira S, Campbell C, Facey M, Baxter NN. What young women with breast cancer get versus what they want in online information and social media supports. J Adolesc Young Adult Oncol 2019 Jun;8(3):320-328. [doi: 10.1089/jayao.2018.0125] [Medline: 30648917]
5. Kemp E, Koczwara B, Butow P, Turner J, Girgis A, Schofield P, et al. Online information and support needs of women with advanced breast cancer: a qualitative analysis. Support Care Cancer 2018 Oct;26(10):3489-3496. [doi: 10.1007/s00520-018-4206-1] [Medline: 29693203]
6. On J, Park H, Song T. Sentiment analysis of social media on childhood vaccination: development of an ontology. J Med Internet Res 2019 Jun 07;21(6):e13456 [FREE Full text] [doi: 10.2196/13456] [Medline: 31199290]
7. Lu Y, Zhang P, Liu J, Li J, Deng S. Health-related hot topic detection in online communities using text clustering. PLoS One 2013 Jan;8(2):e56221 [FREE Full text] [doi: 10.1371/journal.pone.0056221] [Medline: 23457530]
8. Chen AT. Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient Educ Couns 2012 May;87(2):250-257. [doi: 10.1016/j.pec.2011.08.017] [Medline: 21930359]
9. Freedman RA, Viswanath K, Vaz-Luis I, Keating NL. Learning from social media: utilizing advanced data extraction techniques to understand barriers to breast cancer treatment. Breast Cancer Res Treat 2016 Jul;158(2):395-405. [doi: 10.1007/s10549-016-3872-2] [Medline: 27339067]

10. Alpers GW, Winzelberg AJ, Classen C, Roberts H, Dev P, Koopman C, et al. Evaluation of computerized text analysis in an internet breast cancer support group. Computers in Human Behavior 2005 Mar;21(2):361-376. [doi: 10.1016/j.chb.2004.02.008]
11. De Silva D, Ranasinghe W, Bandaragoda T, Adikari A, Mills N, Iddamalgoda L, et al. Machine learning to support social media empowered patients in cancer care and cancer treatment decisions. PLoS One 2018 Oct;13(10):e0205855 [FREE Full text] [doi: 10.1371/journal.pone.0205855] [Medline: 30335805]
12. Bandaragoda T, Ranasinghe W, Adikari A, de Silva D, Lawrentschuk N, Alahakoon D, et al. The Patient-Reported Information Multidimensional Exploration (PRIME) framework for investigating emotions and other factors of prostate cancer patients with low intermediate risk based on online cancer support group discussions. Ann Surg Oncol 2018 Feb 21;25(6):1737-1745. [doi: 10.1245/s10434-018-6372-2]
13. Kaur P, Khamparia A. Diagnosis of liver cancer ontology using SPARQL. International Journal of Applied Engineering Research 2015;10(69):15-18 [FREE Full text]
14. Jusoh F, Ibrahim R, Othman M, Omar N. Development of breast cancer ontology based on hybrid approach. International Journal of Innovative Computing 2013;3(1):15-22. [doi: 10.11113/ijic.v3n1.32]
15. Mahmoodi SA, Mirzaie K, Mahmoudi SM. A new algorithm to extract hidden rules of gastric cancer data based on ontology. Springerplus 2016;5:312 [FREE Full text] [doi: 10.1186/s40064-016-1943-9] [Medline: 27066344]
16. Noy N, McGuinness D. Ontology development 101: A guide to creating your first ontology. Stanford University. 2001. URL: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf [accessed 2019-12-16]
17. Jung H, Park H, Song T. Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. J Med Internet Res 2017 Jul 24;19(7):e259. [doi: 10.2196/jmir.7452]
18. Sentiment Ontology for Social Web. Telecommunications Technology Association. 2013. URL: https://committee.tta.or.kr/summary/standard_view.jsp?publish_year=2013&section=1&pk_num=TTAK.KO-10.0639/R1&nowSu=38&section_code= [accessed 2019-12-16]
19. Comprehensive cancer information. National Cancer Institute. URL: https://www.cancer.gov/ [accessed 2019-12-16]
20. Information and resources about for cancer. American Cancer Society. URL: https://www.cancer.org/ [accessed 2019-12-16]
21. Together we will beat cancer. Cancer Research UK. URL: https://www.cancerresearchuk.org [accessed 2019-12-16]
22. National cancer information center. National Cancer Center. URL: https://www.cancer.go.kr/ [accessed 2019-12-16]
23. National health information portal. Korea Centers for Disease Control and Prevention. URL: https://health.cdc.go.kr [accessed 2019-12-19]
24. Hlomani H, Stacey D. Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: a survey. Semantic Web Journal 2014;1(5):1-11 [FREE Full text]
25. Kehagias D, Papadimitriou I, Hois J, Tzovaras D, Bateman J. A methodological approach for ontology evaluation and refinement. In: ASK-IT Final Conference. 2008 Presented at: ASK-IT Final Conference; June, 2008; Nuremburg, Germany p. 1-13.
26. Buis LR, Whitten P. Comparison of social support content within online communities for high- and low-survival-rate cancers. Comput Inform Nurs 2011 Aug;29(8):461-467. [doi: 10.1097/NCN.0b013e318214093b] [Medline: 21876394]
27. Kim G, Hur H. Information needs on patients with cancer in Korea. The Journal of Korean Academic Society of Adult Nursing 2002;14(1):135-143 [FREE Full text]

Abbreviations

NLP: natural language processing

Edited by G Eysenbach; submitted 17.03.20; peer-reviewed by W Erdley, J Brixey, M Christiansen; comments to author 12.06.20; revised version received 29.07.20; accepted 11.11.20; published 07.12.20.

Please cite as:
Lee J, Park HA, Park SK, Song TM
Using Social Media Data to Understand Consumers' Information Needs and Emotions Regarding Cancer: Ontology-Based Data Analysis Study
J Med Internet Res 2020;22(12):e18767
URL: http://www.jmir.org/2020/12/e18767/
doi:10.2196/18767
PMID:33284127

Original Paper

Facebook-Based Social Marketing to Reduce Smoking in Australia's First Nations Communities: An Analysis of Reach, Shares, and Likes

Marita Hefler[1], BAsianStudies, GradDipIntSocDev, MPH, PhD; Vicki Kerrigan[1], BA, MIntl&ComnDev; Anne Grunseit[2], BA, PhD; Becky Freeman[2], BSc, MSc, PhD; James Kite[2], BA, MPH, PhD; David P Thomas[1], MBBS, MMedSc, PhD, FAFPHM

[1]Menzies School of Health Research, Charles Darwin University, Casuarina, Australia

[2]Prevention Research Collaboration, Sydney School of Public Health, The University of Sydney, Sydney, Australia

Corresponding Author:
Marita Hefler, BAsianStudies, GradDipIntSocDev, MPH, PhD
Menzies School of Health Research
Charles Darwin University
PO Box 41096
Casuarina, 0810
Australia
Phone: 61 889468517
Email: marita.hefler@menzies.edu.au

Abstract

Background: Facebook is widely used by Australia's First Nations people and has significant potential to promote health. However, evidence-based guidelines for its use in health promotion are lacking. Smoking prevalence among Australia's First Nations people is nearly 3 times higher than other Australians. Locally designed programs in Aboriginal Community Controlled Health Services (ACCHOs) to reduce smoking often use Facebook.

Objective: This study reports on an analysis of the reach and engagement of Facebook posts with smoking prevention and cessation messages posted by ACCHOs in the Northern Territory, Australia.

Methods: Each service posted tobacco control content at least weekly for approximately 6 months. Posts were coded for the following variables: service posted, tailored First Nations Australian content, local or nonlocally produced content, video or nonvideo, communication technique, and emotional appeal. The overall reach, shares, and reactions were calculated.

Results: Compared with posts developed by the health services, posts with content created by other sources had greater reach (adjusted incident rate ratio [IRR] 1.92, 95% CI 1.03-3.59). Similarly, reactions to posts (IRR 1.89, 95% CI 1.40-2.56) and shared posts (IRR 2.17, 95% CI 1.31-3.61) with content created by other sources also had more reactions, after controlling for reach, as did posts with local First Nations content compared with posts with no First Nations content (IRR 1.71, 95% CI 1.21-2.34).

Conclusions: Facebook posts with nonlocally produced content can be an important component of a social media campaign run by local health organizations. With the exception of nonlocally produced content, we did not find a definitive set of characteristics that were clearly associated with reach, shares, and reactions. Beyond reach, shares, and likes, further research is needed to understand the extent that social media content can influence health behavior.

(J Med Internet Res 2020;22(12):e16927) doi:10.2196/16927

KEYWORDS

social media; tobacco; Australia; indigenous peoples; smoking; health promotion

Introduction

Social Media and Health

Social media are an essential component of public health, as a tool for health communication, an avenue for consumers seeking information, and an environment that shapes health, analogous to the influence of the built environment [1]. Although social media have fundamentally altered health promotion [2], considerable uncertainty remains regarding how health services can use them effectively [3,4]. The rapid change in the social media landscape challenges the capacity of health promoters to develop a robust evidence base for what works; as a result, there are no clear guidelines for population-based social media strategies to promote health. Numerous reviews have concluded that social media have the potential to positively affect health but have found either mixed [5,6] or no definitive evidence [7-11] of their impact on health behavior, even as research in the field has significantly increased in recent years.

Consumer benefits of using social media for health information include social and emotional support and connectedness from peer-to-peer interactions [12-14]. Similarly, social media are an opportunity for health services to engage in building social capital within their communities [15] and a potentially cost-effective way to achieve significant reach for social marketing and health promotion [3]. Social media may also be effective for engaging specific groups who experience disadvantage and, therefore, contribute to reducing health inequity, although more research is needed [10].

Social marketing and mass media are an important and cost-effective strategy for reducing smoking [16-19]. Social media are an increasingly integral component of mass media strategies, and Australian research suggests that social media alone are more cost-effective than television advertising for tobacco control social marketing [20]. Social media also complement strategies such as smoking cessation services [21]. Facebook can also be effective for connecting with otherwise hard-to-reach smokers [22], including with people who are not social media users when people show content that appears on their Facebook feed to people who do not have an account [15].

Worldwide, Facebook is the largest social media platform, with nearly 2.8 billion active users [23]. In Australia, it is the most widely and frequently used platform, with 15 million active users (equivalent to 60% of the population) and 50% of the population using it daily [24]. Facebook is a comparatively inexpensive way for health services to communicate with the public [25], allowing for a spectrum of engagement from low to high [26]. Unlike many other platforms, Facebook is used by all ages, including older adults, for whom it can have both social and health benefits [13,14,27]. Studies examining the characteristics of posts that generate engagement have found diverse results; some have found videos to be effective [28], others have found photos to be effective [29], and some have identified that paid boosts and page promotions stimulate engagement [26,30] and that the time of day or day of week and organic versus paid posts are also factors that determine both reach and engagement [30]. The Facebook algorithm appears to increasingly prioritize posts from personal contacts rather than unpaid posts from organizational or business pages [31]. In this context, health services need to understand how they can use Facebook to reach key target groups and track both reach and engagement.

Social Media and Australia’s First Nations People

Although data are limited, social media—and specifically Facebook—use appears to be higher among Australia’s First Nations people than other Australians [32-34]. Facebook has been found to be a supportive environment which is used by Australia’s First Nations people to share health information [35,36]. The high use of social media by Australia’s First Nations peoples is not limited to urban areas; particularly in remote communities, Australia’s First Nations people are likely to be *mobile only* users in relation to communication, with social media access often occurring through shared devices [37].

Unsurprisingly, the lack of clear evidence about how to use social media effectively to improve health is reflected in reviews about how social media can be used to improve the health of Australia’s First Nations people [36,38]. However, research is promising that social media can help with promoting healthy behaviors [39]. In Australia, the use of social media by First Nations people has some specific characteristics, including engagement between youth and older adults, which promotes intergenerational connection [33]. Social media is therefore an appropriate channel for communicating messages to Australia’s First Nations people as part of social marketing and mass media campaigns.

Australia’s First Nations People and Smoking

Nationally, smoking prevalence among Australia’s First Nations people is nearly 3 times higher than that among non-Indigenous Australians [40]. To address this, the national Tackling Indigenous Smoking Program is funded by the Australian government till June 30, 2022. It is implemented through 37 Aboriginal Community Controlled Health Services (ACCHOs) to deliver locally designed programs to reduce smoking in their communities [41] and includes the use of mass media, both legacy (television, radio, and newspaper) and social media. These local health services therefore play an essential role in disseminating information to localize and make relevant key tobacco control messages and generate awareness of cessation support for smokers.

Study Objective

This study reports on an analysis of Facebook post reach and engagement with smoking prevention and cessation messages posted by ACCHOs in the Northern Territory, Australia.

A note regarding terminology: the original inhabitants of what is now known as Australia have diverse cultures, languages, and kinship structures. Throughout this paper, the term Australia’s First Nations is used, unless referring specifically to service or program names that use Indigenous or Aboriginal in the title.

Methods

Setting

In 2016, the Northern Territory had an estimated population of 245,048 people [42]. The Northern Territory has both the highest proportion of First Nations people (23.8%) [42] and the highest smoking prevalence compared with any other Australian jurisdiction. Among people aged 18 years or older, the daily smoking prevalence in the Northern Territory in 2017 to 2018 was 19.6%, compared with 13.8% of all Australians [43]. In 2014, the daily smoking prevalence among First Nations people in the Northern Territory was 45% [44].

This study was undertaken with 3 ACCHOs in the Northern Territory, located in a mix of urban and remote locations. It was designed to meet an identified need for services to have a better, context-specific understanding of what content works well with their communities. Implementation was participatory, flexible, and pragmatic, based on real-world circumstances for, and decision making by, each service (rather than imposing a set approach to selecting and posting content). All the services received funding from the Tackling Indigenous Smoking Program. The approach and factors explored in this study were based on guidance from the services that were partners in this project and feedback from other services that receive Tackling Indigenous Smoking Program funding. One participating service (hereafter referred to as service 3) had an organizational Facebook page before the commencement of the study; the other 2 started their Facebook pages during the study. Data were collected from the time of the first tobacco-related post during the study period. Each service posted a mix of tobacco control content, other health information, and general information during the study period. Data collected and analyzed for this study only included tobacco-related posts. One health service paid to boost the reach of 2 posts. With only 2 posts, it was not possible to assess the impact of these posts on reach.

One health service that started its Facebook page during the study had an in-house specialist communications manager who was responsible for coordinating and managing posts for this study, as part of managing social media communications for the service. The other 2 services received support and mentoring from VK (project manager with the research institute) throughout the study period, including technical assistance to develop content, support for crafting messages, and risk management for negative public comments. This technical support was part of the overall project design, as requested by the partner services. VK has 2 decades of experience as a journalist, including administration of local Facebook accounts for a national media outlet. The support provided by VK was part of capacity building and reciprocity, which was embedded in the project approach, in line with Australian national guidelines for ethical conduct in Aboriginal and Torres Strait Islander health research [45]. The project was approved by the Northern Territory Department of Health and Menzies School of Health Research Human Research Ethics Committee and the Central Australian Human Research Ethics Committee, although ethics approval was not required for this component.

Data Collection

Facebook analytic data (available to page administrators) were collected for the duration of each health service's participation in the project. Data were collected 7 to 14 days after each post. The difference in time between posting and data collection is unlikely to have affected the overall reach, as most reach and interactions with Facebook posts are known to be achieved within 24 hours of posting [46]. The study period and timeframe were different for each service to take into account operational priorities, staff capacity, and holiday close down periods.

Outcome Variables

We collected the following outcome measures for each post:

- Overall reach: the total number of unique users exposed to the original post and shares, as defined by Facebook. Facebook insights do not provide a breakdown of exposure to the original post and shares.
- Total shares: includes shares of the original post and on-shares (shares of the shared post). This category also includes shares of posts that the health service posted on its own page and then shared to the page of another organization.
- Reactions to original post: all reactions (like, love, laugh, angry, sad, and wow) and comments on a post by the health service. Does not include shares of the post or reactions to on-shares.
- Reactions to shared posts: all reactions (like, love, laugh, angry, sad, and wow) and comments to posts from the health service that were shared by others.

Reactions to original posts and shared posts were analyzed separately to provide insight into differences in the magnitude of reactions to original posts compared with shared posts. Furthermore, it was not always possible to determine what was included in shared posts—for example, comments by the health services on the original post may not have been included or people may have added their own comments when sharing posts that were not visible to the research team.

Independent Variables

We adapted previous coding frames [28,47] for post content to take into account the smaller volume of posts analyzed in this study and ensure that codes were locally relevant. Coding criteria were determined by agreement between 3 authors (VK, MH, and DT), with potential discrepancies resolved at this stage. Codes were then assigned by 1 author (VK; Textbox 1), except for *emotional appeal* (see definition in Textbox 1).

Textbox 1. Coding frame for post content (item and definition).

- Who posted: Health service 1, 2, or 3
- First Nations (Australia) tailoring:
    - Non–First Nations: General tobacco control messages, whether Australian or international, not featuring any First Nations people, words or images
    - First Nations Australia—general: Australian content, tailored to First Nations people, but not specific to the health service location. Includes content from locations in the Northern Territory, but outside the catchment area or footprint of the health service, and content developed by other services in the project
    - First Nations Australia—local: Content which has features that are localized to health service area. Includes First Nations Australia general content with local branding features such as local health service logo or translated into language
- Content origin:
    - Original: Content developed by the health service which posted the content
    - Other sources: Content created by a source other than the health service, even if localized to the health service (eg, Australian Government Department of Health “Don’t make smokes your story” campaign images translated into the local language and/or with a logo of a local health service added)
- Post type: Includes or does not include video
- Communication techniques:
    - Other orgs tagged—no or yes
    - Hashtags used—yes or no
- Emotional appeal: As this is influenced by cultural background, posts from each health service were coded by 2 community members living in the catchment area of each service. Coders were requested to code each post as negative, positive, or neutral. Coding was based on their perceptions of the post; the intended emotional appeal of the content was not known by the researchers or coders. If coders perceived the post as having no emotional elements or both positive and negative elements, they were asked to code it as neutral. Posts that were coded differently by each community member were combined with those that were coded as neutral by both. As no posts were coded as negative by the coders, 2 final categories were used for analysis: (1) positive and (2) neutral or disagreement (contested) between coders

First, we calculated the number of posts, shares, reactions, and median reach per post over the independent variables. Second, we conducted a series of (separate) bivariate negative binomial regressions (the data were overdispersed) for each outcome (reach, shares, reactions to original post, and reaction to shared posts) over each of the 6 explanatory variables, and then multivariate regressions including all of the explanatory variables. The reference category for health service was health service 1. For First Nations tailoring, it was non–First Nations content; for content origin, it was original content; for post type, it was nonvideo; and for whether other organizations were tagged or hashtags were used, the reference categories were not tagged and no hashtags, respectively, and neutral or mixed was the reference category for emotional appeal.

Estimated coefficients for multicategory variables (health service and First Nations tailoring) were tested jointly postestimation for statistical significance of the variable using Wald tests. The analyses were initially conducted without an offset and then controlling for users’ exposure to the post by including reach as an offset variable to estimate engagement while accounting for the number of people to whom each post was delivered [48]. For the unadjusted analyses, the outcome (incident rate ratio [IRR]) indicates the multiplicative change in outcome compared with the reference category for that explanatory variable. For the adjusted analyses, the relationship between explanatory variables and engagement becomes the change in the outcome per person. The threshold for statistical significance was set at *P*=.05.

Results

Service 1 participated for 23 weeks from August 2017, service 2 for 19 weeks from July 2017 to March 2018, and service 3 for 19 weeks from May to December 2017. We coded 92 posts from the study period (Table 1).

Table 1. Overall reach of each type of post (both of the original post and when shared).

Variable	Posts, n	Overall reach						
		Reach per post, median	IRR[a]	95% CI	*P* value[b]	Adjusted IRR[c]	95% CI	*P* value[b]
Total	92	248	N/A[d]	N/A	N/A	N/A	N/A	N/A
Health service					<.001			<.001
1	33	177	1	N/A		1	N/A	
2	25	114	0.94	0.56-1.59		1.11	0.61-2.04	
3	34	1432	5.20	3.22-8.39		6.18	3.33-11.47	
First Nations (Australia) content					.11			.20
No	20	448	1	N/A		1	N/A	
Yes, not local	14	189	0.40	0.17-0.91		0.94	0.46-1.96	
Yes, local	58	248	0.69	0.37-1.28		0.61	0.33-1.11	
Content source					.008			.04
Original content	67	192	1	N/A		1	N/A	
Other sources	25	1014	2.10	1.21-3.60		1.92	1.03-3.59	
Video					.02			.72
No	70	244	1	N/A		1	N/A	
Yes	22	762	1.95	1.10-3.46		0.90	0.49-1.64	
Other organizations tagged					.25			.34
No	83	247	1	N/A		1	N/A	
Yes	9	488	1.63	0.71-3.78		1.48	0.66-3.33	
Hashtagged					.33			.39
No	14	312	1	N/A		1	N/A	
Yes	78	248	1.42	0.71-2.84		1.37	0.67-2.80	
Tone of content					.19			.02
Neutral or disagreement between coders[e]	48	244	1	N/A		1	N/A	
Positive	44	294	1.39	0.85-2.29		0.51	0.29-0.90	

[a] IRR: incident rate ratio.

[b] *P* value calculated for whole variable using chi-square test.

[c] Adjusted IRR for all other variables describing types of posts using multiple negative binomial regression.

[d] N/A: not applicable.

[e] Includes posts where all coded as neutral content and posts where coders could not agree on them being positive, neutral, or negative. No posts were coded as negative in tone or content by all coders.

Most posts (58/92, 63%) featured local First Nations content (rather than either no Australian First Nations content or content that was not local) and original content created by the health service (67/92, 73% posts; rather than content from other sources). There were no posts that featured content that was perceived as negative by all coders, such as intended negative emotional arousal featured in fear appeal TV campaigns; there was disagreement between coders on 37 posts, and the remainder were coded by all coders as either positive (44 posts) or neutral (11 posts). As noted in Textbox 1, neutral and posts with coder disagreement were treated as 1 category. Only 22 of 92 posts (24%) included videos, and although most (78/92, 85%) included hashtags, fewer than 10% tagged other organizations. One service paid to boost reach for 2 posts; all other posts (90/92, 97%) were organic (nonpaid page posts).

The median overall reach of each post was 248 unique users. Posts reached more people if they were from health service 3, which had an established Facebook page preproject, if the featured content was from other sources (rather than original content created by the health service) or if the posts contained videos. Although there was no association with reach in the bivariate analysis, once adjusted for other post features, having positive content was associated with less reach (compared with neutral or contested content). In relation to the service that boosted the reach of 2 posts, we conducted a sensitivity analysis of our results by removing the 2 paid posts; the same variables had significant associations in all tables.

The 92 posts were shared 352 times (median 1 share per post; IQR 0-4; range 0-45; Table 2). Posts from health services 2 and 3 were shared more than those from health service 1. Other characteristics were not associated with the sharing of posts (Multimedia Appendix 1).

Table 2. Total shares of each type of post, with and without controlling for reach.

Variable	Shares, n	Shares, Median	IRR[a]	95% CI	*P* value[b]	IRR (controlling for reach)[c]	95% CI	*P* value[b]
Total	352	1	N/A[d]	N/A	N/A	N/A	N/A	N/A
Health service					<.001			.005
1	41	0	1	N/A		1	N/A	
2	61	1	1.96	0.85-4.56		2.90	1.43-5.88	
3	250	6	5.92	2.76-12.67		2.77	1.47-5.21	
First Nations (Australia) content					.22			.45
No	127	2	1	N/A		1	N/A	
Yes, not local	48	0	0.54	0.17-1.74		0.80	0.33-1.94	
Yes, local	177	1	0.48	0.20-1.14		0.66	0.34-1.27	
Content source					.37			.95
Original content	229	1	1	N/A		1	N/A	
Other sources	123	4	1.44	0.65-3.20		1.02	0.56-1.88	
Video					.20			.23
No	229	1	1	N/A		1	N/A	
Yes	123	4	1.71	0.75-3.90		1.47	0.78-2.75	
Other organizations tagged					.82			.87
No	313	1	1	N/A		1	N/A	
Yes	39	4	1.15	0.34-3.83		1.08	0.44-2.65	
Hashtagged					.26			.58
No	32	1	1	N/A		1	N/A	
Yes	320	1	1.79	0.65-4.96		1.25	0.56-2.78	
Tone of content					.12			.37
Neutral or disagreement between coders	135	1	1	N/A		1	N/A	
Positive	217	1	1.75	0.86-3.56		1.29	0.74-2.24	

[a] IRR: incident rate ratio.

[b] *P* values calculated for whole variables using chi-square test. We controlled for reach by offsetting the negative binomial regression model by the reach of the original posts.

[c] We controlled for reach by offsetting the negative binomial regression model by the reach of the original posts.

[d] N/A: not applicable.

There were 1099 reactions to the original posts (median 7 reactions per post; IQR 3-13; range 0-94; Table 3). In the bivariate analyses, posts from health service 3 (compared with health service 1) with content from other sources (rather than original content created by the health service) featuring video and with other organizations tagged attracted more reactions (likes or comments). After controlling for the reach of the posts, however, the differences between health services became insignificant, and posts with local Australian First Nations content attracted more reactions per person reached (compared with posts with no Australian First Nations content).

After adjusting for all variables, posts with content from other sources (compared with posts with original content) attracted significantly more reactions (IRR 1.57, 95% CI 1.08-2.27; *P*=.02; Multimedia Appendix 2).

There were 990 reactions to the shared posts (median zero reactions per shared post; IQR 0-10; range 0-97; Table 4).

Table 3. Reactions to each type of original post, before and after controlling for reach.

Variable	Reactions, n	Reactions, Median	IRR[a]	95% CI	*P* value[b]	IRR (controlling for reach)[c]	95% CI	*P* value[b]
Total	1099	7	N/A[d]	N/A	N/A	N/A	N/A	N/A
Health service					<.001			.25
1	334	4	1	N/A		1	N/A	
2	162	5	0.64	0.38-1.08		1.30	0.88-1.94	
3	603	10	1.75	1.09-2.81		0.95	0.67-1.35	
First Nations (Australia) content					.03			.003
No	182	9	1	N/A		1	N/A	
Yes, not local	92	5	0.72	0.35-1.48		0.96	0.59-1.58	
Yes, local	825	6	1.56	0.93-2.64		1.71	1.19-2.46	
Content source					<.001			<.001
Original content	517	4	1	N/A		1	N/A	
Other sources	582	19	3.02	1.98-4.60		1.89	1.40-2.56	
Video					.002			.002
No	661	6	1	N/A		1	N/A	
Yes	438	11	2.11	1.31-3.40		1.68	1.21-2.34	
Other organizations tagged					.03			.02
No	890	6	1	N/A		1	N/A	
Yes	209	19	2.17	1.08-4.34		1.74	1.09-2.78	
Hashtagged					.83			.31
No	158	7	1	N/A		1	N/A	
Yes	941	8	1.07	0.59-1.95		0.80	0.53-1.23	
Tone of content					.97			.25
Neutral or disagreement between coders	571	6	1	N/A		1	N/A	
Positive	528	8	1.01	0.65-1.55		0.84	0.62-1.13	

[a] IRR: incident rate ratio.

[b] *P* values calculated for whole variables using the chi-square test. We controlled for reach by offsetting the negative binomial regression model by the reach of the original posts.

[c] We controlled for reach by offsetting the negative binomial regression model by the reach of the original posts.

[d] N/A: not applicable.

Table 4. Reactions to shared posts from each type of original post, before and after controlling for reach of shared posts.

Variable	Reactions, n	Reactions, median	IRR[a]	95% CI	*P* value[b]	IRR (controlling for reach)[c]	95% CI	*P* value[b]
Total	990	0	N/A[d]	N/A	N/A	N/A	N/A	N/A
Health service					.003			.02
1	74	0	1	N/A		1	N/A	
2	215	0	3.84	1.13-13.02		1.99	0.99-3.98	
3	701	10	9.19	2.98-28.33		0.87	0.47-1.61	
First Nations (Australia) content					.78			.29
No	231	1	1	N/A		1	N/A	
Yes—not local	92	0	0.57	0.10-3.25		1.54	0.62-3.84	
Yes—local	667	0	1.00	0.27-3.62		1.69	0.90-3.18	
Content source					.02			.003
Original content	399	0	1	N/A		1	N/A	
Other sources	591	10	3.97	1.31-12.05		2.17	1.31-3.61	
Video					.21			.57
No	589	0	1	N/A		1	N/A	
Yes	401	8	2.17	0.65-7.22		1.18	0.67-2.10	
Other organizations tagged					.29			.11
No	777	0	1	N/A		1	N/A	
Yes	213	34	2.53	0.45-14.27		1.86	0.87-3.95	
Hashtagged					.58			.002
No	211	1	1	N/A		1	N/A	
Yes	779	0	0.66	0.16-2.81		0.36	0.19-0.68	
Tone of content					.52			.27
Neutral or disagreement between coders	433	0	1	N/A		1	N/A	
Positive	557	1	1.40	0.50-3.98		0.75	0.44-1.26	

[a] IRR: incident rate ratio.

[b] *P* values calculated for whole variables using the chi-square test. We controlled for reach by offsetting the negative binomial regression model by the reach of the shared posts.

[c] We controlled for reach by offsetting the negative binomial regression model by the reach of the original posts.

[d] N/A: not applicable.

Shared posts from original posts with content from nonlocal sources (rather than original content created by the health service) attracted more reactions, which persisted after controlling for reach and other post characteristics (IRR 1.57, 95% CI 1.08-2.27; *P*=.02; Multimedia Appendix 3). Shared posts from health services 2 and 3 (compared with health service 1) had more reactions, but the magnitude of this association decreased after adjusting for reach and became nonsignificant once other characteristics were included in the model (Multimedia Appendix 3).

Discussion

We found that reach and sharing of posts was largely associated with the health service posting the content. This is unsurprising, given that only 1 service had an established Facebook presence before the study started. Otherwise, we found mixed evidence about other factors, including the use of First Nations–specific content, positive emotional content, video, tagging, and use of hashtags.

A counter-intuitive finding was that reactions to both original and shared posts were associated with content from other sources, not original content. A potential explanation for this could be the lower production quality of locally generated content. However, in this study, many of the posts with locally produced content were of high production quality, including being designed to be of appropriate length and tone for use on social media. Much of the locally generated content was developed and produced with the assistance of a former journalist and author VK, who has extensive professional experience in this area. Given that production quality is unlikely to be an issue, the results could be explained by 2 main factors: novelty and relevance.

Content from other sources may have had a higher novelty value than the local content; other research has shown that novelty can be an important factor in generating engagement [49]. This result may also reflect an effort to be particularly selective about choosing content from other sources, particularly when it contained nonlocal content, which is perceived to be specifically relevant to the local community. Our previous research [15,50] and research by others [44] has found that sharing posts with content obtained from other sources but with a personalized message and posted by a close and trusted family member or friend can be perceived as localized, particularly if the content is highly resonant for the local context (for example, a post that conveys a commonly encountered scenario). Together, these factors may explain why content from other sources generated greater engagement. Further research is required to determine which of these factors explains our results and whether content creators can leverage them to boost engagement with their content.

The more mixed results for First Nations–specific content, both local and nonlocal, may be due to less selectivity being applied when choosing which content to post. It is also possible that, despite local branding and language, and high production values, the generic nature of the Australian government posts had a lower novelty value than other posts. It may also have been perceived as government content, which we have previously found is less trusted than content that is genuinely local [15].

Another unexpected finding was that positive content was associated with lower reach than posts that were coded as neutral and/or contested. Australia’s First Nations people have called for change from deficit discourse to strengths-based messages and research approaches [51]. In our previous research, participants selected positive content for sharing, and the importance of contributing to a supportive environment was highlighted [15,34,50]. However, those studies explored decision making for sharing posts on personal pages, which may differ from the perceptions of posts by organizations. An evaluation of an Australian state government health promotion campaign’s Facebook page found that users expressed a preference for positive content [49], but such content was not associated with higher engagement [30]. The lower reach achieved by positive content in our study may reflect the fact that the coding of most of the remaining posts was contested between the 2 coders or was coded as neutral because it included both positive and negative elements. It may be that unambiguously positive posts were perceived as less interesting or noteworthy than the posts with more ambiguous emotional content. Neutral and contested posts may have contained elements more likely to challenge and stimulate interpersonal discussion. This is important, as interpersonal communication stimulated by exposure to mass media antismoking campaigns is a factor in quitting intentions and behavior [52,53]. In previous research, we have also found evidence that this occurs in response to Facebook posts from friends and family about smoking [15]. We would not recommend avoiding posts with a positive tone based on these new data. Given that people tend to state a preference for positive content, it is likely that this is important for generating a positive connection to a page and receptiveness to messages posted. Therefore, positive content is likely to be complementary to a negative tone or mixed posts that may generate greater engagement.

Although our results were inconclusive, we were able to identify some areas for further research. First, as only 2 posts in our study received paid boosts, we did not perform a separate analysis of these factors. Previous research has found that the use of paid posts can significantly boost reach; however, it has also shown that organic (nonpaid) reach is associated with generating greater page engagement per person reached [30], suggesting that high quality, relevant, and appealing content is still essential if a Facebook campaign is to be successful. Other research has found that page administrators can leverage contextual factors that promote user engagement with pages (eg, the perceived trustworthiness of page, user patterns of use), but these factors cannot be controlled by page administrators [49]. Given the unique role of ACCHOs in delivering health services to First Nations Australians [54,55], posts by ACCHOs may generate engagement because they are seen as a worthy initiative.

The strengths of this study include that the content posted was selected and posted by the partner ACCHOs, based on what they perceived as relevant for their communities. Posting was integrated into routine work, ensuring that the approach was pragmatic and sustainable beyond this study. There were also a number of notable limitations. These included the fact that we did not perform separate analyses for different types of reactions, such as likes and comments, due to the small number of posts and that we were unable to extend our analysis beyond metrics of lower-level engagement, according to a spectrum of engagement outcomes as has been performed in other research [26]. Furthermore, we examined only postlevel data, not the impact of posts on page-level data. In addition, the sparsely populated geographical area, sample size (number of posts), and short duration limit the generalizability of our results. Finally, data from end users, which could have provided insights into the types of content that would be most likely to generate engagement, are not included in this paper.

Conclusions

Overall, we did not find a definitive set of characteristics that were clearly associated with reach and engagement. The exception was that content from other sources was associated with higher engagement than original content, which was an unexpected finding. However, it is important to note that our results do not suggest that community-based, localized content should be avoided. Similarly, although posts perceived by community members to have a positive tone generated less engagement, we do not suggest that these should be avoided.

Rather, positive tone posts could be complementary to posts with characteristics that are likely to generate greater interaction. One of the key appeals of social media, specifically Facebook, is its ability to reach relatively small populations and to tailor content accordingly. For Australia's First Nations people, it can help with building social capital and connection. This study contributes to an understanding of the role of localized, community-based social media efforts and how these can be implemented. However, additional research is needed to understand the extent to which social media content can influence behavior.

Acknowledgments

The authors would like to thank the project partners Danila Dilba Health Service, Miwatj Health Aboriginal Corporation, Central Australian Aboriginal Congress (Congress), Aboriginal Medical Services Alliance Northern Territory (AMSANT), and all project participants.

This study was funded by the National Health and Medical Research Council, Australia, project grant number 1098308. MH is supported by a National Heart Foundation Australia Postdoctoral Fellowship award number 101938. BF is supported by a National Health and Medical Research Council early career fellowship APP1089403. The funding bodies had no role in the design of the study, data collection, analysis, and interpretation, or in writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Total shares of post controlling for reach, before and after adjusting for all other variables.
[DOCX File , 25 KB - jmir_v22i12e16927_app1.docx]

Multimedia Appendix 2
Total reactions to original posts controlling for reach, before and after adjusting for all other variables.
[DOCX File , 16 KB - jmir_v22i12e16927_app2.docx]

Multimedia Appendix 3
Total reactions to shared posts controlling for reach, before and after adjusting for all other variables.
[DOCX File , 17 KB - jmir_v22i12e16927_app3.docx]

References

1. Abroms LC. Public health in the era of social media. Am J Public Health 2019 Feb;109(S2):S130-S131. [doi: 10.2105/ajph.2018.304947]
2. Kickbusch I. Health promotion 4.0. Health Promot Int 2019 Apr 1;34(2):179-181. [doi: 10.1093/heapro/daz022] [Medline: 30939200]
3. Jane M, Hagger M, Foster J, Ho S, Pal S. Social media for health promotion and weight management: a critical debate. BMC Public Health 2018 Jul 28;18(1):932 [FREE Full text] [doi: 10.1186/s12889-018-5837-3] [Medline: 30055592]
4. Partridge SR, Gallagher P, Freeman B, Gallagher R. Facebook groups for the management of chronic diseases. J Med Internet Res 2018 Jan 17;20(1):e21 [FREE Full text] [doi: 10.2196/jmir.7558] [Medline: 29343460]
5. Hunter RF, de la Haye K, Murray JM, Badham J, Valente TW, Clarke M, et al. Social network interventions for health behaviours and outcomes: a systematic review and meta-analysis. PLoS Med 2019 Sep;16(9):e1002890 [FREE Full text] [doi: 10.1371/journal.pmed.1002890] [Medline: 31479454]
6. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
7. Johns D, Langley TS, Lewis S. Use of social media for the delivery of health promotion on smoking, nutrition, and physical activity: a systematic review. Lancet 2017 Nov;390:S49. [doi: 10.1016/s0140-6736(17)32984-7]
8. Giustini D, Ali SM, Fraser M, Kamel Boulos MN. Effective uses of social media in public health and medicine: a systematic review of systematic reviews. Online J Public Health Inform 2018;10(2):e215 [FREE Full text] [doi: 10.5210/ojphi.v10i2.8270] [Medline: 30349633]
9. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 5;16(Suppl 5):290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]

10. Welch V, Petkovic J, Pardo Pardo J, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75 [FREE Full text] [doi: 10.24095/hpcdp.36.4.01] [Medline: 27077792]
11. Moorhead A, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
12. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283 [FREE Full text] [doi: 10.1111/hir.12192] [Medline: 29045011]
13. Yu RP, Ellison NB, Lampe C. Facebook use and its role in shaping access to social benefits among older adults. J Broadcast Elect Media 2018 Jan 30;62(1):71-90. [doi: 10.1080/08838151.2017.1402905]
14. Sinclair TJ, Grieve R. Facebook as a source of social connectedness in older adults. Comput Hum Behav 2017 Jan;66:363-369. [doi: 10.1016/j.chb.2016.10.003]
15. Kerrigan V, Herdman RM, Thomas DP, Hefler M. 'I still remember your post about buying smokes': a case study of a remote Aboriginal community-controlled health service using Facebook for tobacco control. Aust J Prim Health 2019;25(5):443. [doi: 10.1071/py19008]
16. Kuipers M, Beard E, West R, Brown J. Associations between tobacco control mass media campaign expenditure and smoking prevalence and quitting in England: a time series analysis. Tob Control 2018 Jul;27(4):455-462 [FREE Full text] [doi: 10.1136/tobaccocontrol-2017-053662] [Medline: 28667091]
17. Durkin S, Brennan EB, Wakefield M. Mass media campaigns to promote smoking cessation among adults: an integrative review. Tob Control 2012 Mar;21(2):127-138. [doi: 10.1136/tobaccocontrol-2011-050345] [Medline: 22345235]
18. Brown J, Kotz D, Michie S, Stapleton J, Walmsley M, West R. How effective and cost-effective was the national mass media smoking cessation campaign 'Stoptober'? Drug Alcohol Depend 2014 Feb 1;135:52-58 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.11.003] [Medline: 24322004]
19. Xu X, Alexander RL, Simpson SA, Goates S, Nonnemaker JM, Davis KC, et al. A cost-effectiveness analysis of the first federally funded antismoking campaign. Am J Prev Med 2015 Mar;48(3):318-325 [FREE Full text] [doi: 10.1016/j.amepre.2014.10.011] [Medline: 25498550]
20. Allom V, Jongenelis M, Slevin T, Keightley S, Phillips F, Beasley S, et al. Comparing the cost-effectiveness of campaigns delivered various combinations of television and online media. Front Public Health 2018;6:83 [FREE Full text] [doi: 10.3389/fpubh.2018.00083] [Medline: 29629366]
21. Baskerville NB, Azagba S, Norman C, McKeown K, Brown KS. Effect of a digital social media campaign on young adult smoking cessation. Nicotine Tob Res 2016 Mar;18(3):351-360. [doi: 10.1093/ntr/ntv119] [Medline: 26045252]
22. Bold KW, Hanrahan TH, O'Malley SS, Fucito LM. Exploring the utility of web-based social media advertising to recruit adult heavy-drinking smokers for treatment. J Med Internet Res 2016 May 18;18(5):e107 [FREE Full text] [doi: 10.2196/jmir.5360] [Medline: 27194456]
23. Most Popular Social Networks Worldwide as of July 2019, Ranked by Number of Active Users (in Millions). Statista. 2019. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2019-11-05]
24. Cowling D. Social Media Statistics Australia. Social Media News. 2019. URL: https://www.socialmedianews.com.au/social-media-statistics-australia-july-2019/ [accessed 2019-09-13]
25. Kordzadeh ND, Young DK. Exploring hospitals' use of Facebook: thematic analysis. J Med Internet Res 2018 May 16;20(5):e190 [FREE Full text] [doi: 10.2196/jmir.9549] [Medline: 29769173]
26. Platt T, Platt J, Thiel DB, Kardia SL. Facebook advertising across an engagement spectrum: a case example for public health communication. JMIR Public Health Surveill 2016 May 30;2(1):e27 [FREE Full text] [doi: 10.2196/publichealth.5623] [Medline: 27244774]
27. Chopik WJ. The benefits of social technology use among older adults are mediated by reduced loneliness. Cyberpsychol Behav Soc Netw 2016 Sep;19(9):551-556 [FREE Full text] [doi: 10.1089/cyber.2016.0151] [Medline: 27541746]
28. Kite J, Foley BC, Grunseit AC, Freeman B. Please like me: Facebook and public health communication. PLoS One 2016;11(9):e0162765 [FREE Full text] [doi: 10.1371/journal.pone.0162765] [Medline: 27632172]
29. Theiss S, Burke RM, Cory JL, Fairley TL. Getting beyond impressions: an evaluation of engagement with breast cancer-related Facebook content. Mhealth 2016 Nov;2:- [FREE Full text] [doi: 10.21037/mhealth.2016.10.02] [Medline: 27840816]
30. Kite J, Grunseit A, Li V, Vineburg J, Berton N, Bauman A, et al. Generating engagement on the make healthy normal campaign Facebook page: analysis of Facebook analytics. JMIR Public Health Surveill 2019 Jan 14;5(1):e11132 [FREE Full text] [doi: 10.2196/11132] [Medline: 31344679]
31. Barnhart B. How the Facebook Algorithm Works and Ways to Outsmart It. Sprout Social. 2019. URL: https://sproutsocial.com/insights/facebook-algorithm/ [accessed 2019-09-13]
32. Carlson B. For Aboriginal and Torres Strait Islander people what are the health impacts of social media?2016. Croakey. 2019. URL: https://croakey.org/for-aboriginal-and-torres-strait-islander-people-what-are-the-health-impacts-of-social-media/ [accessed 2019-09-13]

33. Carlson B, Frazer R. Social Media Mob: Being Indigenous Online. Macquarie Park, NSW: Macquarie University; 2018.
34. Hefler M, Kerrigan V, Henryks J, Freeman B, Thomas DP. Social media and health information sharing among Australian Indigenous people. Health Promot Int 2019 Aug 1;34(4):706-715 [FREE Full text] [doi: 10.1093/heapro/day018] [Medline: 29672684]
35. Hefler M, Kerrigan V, Henryks J, Freeman B, Thomas DP. Social media and health information sharing among Australian Indigenous people. Health Promot Int 2019 Aug 1;34(4):706-715 [FREE Full text] [doi: 10.1093/heapro/day018] [Medline: 29672684]
36. Walker T, Palermo CK, Klassen K. Considering the impact of social media on contemporary improvement of Australian aboriginal health: scoping review. JMIR Public Health Surveill 2019 Feb 5;5(1):e11573 [FREE Full text] [doi: 10.2196/11573] [Medline: 30720442]
37. Rennie E, Yunkaporta T, Holcombe-James I. Privacy at the margins| privacy versus relatedness: managing device use in Australia’s remote aboriginal communities. Int J Commun 2018 Mar;12:1291-1309 [FREE Full text]
38. Brusse C, Gardner K, McAullay D, Dowden M. Social media and mobile apps for health promotion in Australian Indigenous populations: scoping review. J Med Internet Res 2014 Dec 10;16(12):e280 [FREE Full text] [doi: 10.2196/jmir.3614] [Medline: 25498835]
39. Berends L, Halliday R. Capacity building and social marketing promotes healthy lifestyle behaviour in an Australian Aboriginal community. Aust J Rural Health 2018 Jul 16:-. [doi: 10.1111/ajr.12421] [Medline: 30011105]
40. Australian Aboriginal and Torres Strait Islander Health Survey: Updated Results, 2012–13. Australian Bureau of Statistics. Canberra; 2014. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/4727.0.55.006 [accessed 2019-09-13]
41. Tackling Indigenous Smoking. Australian Government. 2019. URL: https://www.health.gov.au/initiatives-and-programs/tackling-indigenous-smoking [accessed 2019-09-13]
42. 2016 Census Reveals the Changing Face of the Northern Territory. Australian Bureau of Statistics. 2017. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/mediareleasesbyReleaseDate/C73D7CC81CA1FD2FCA258148000A4067?OpenDocument [accessed 2019-09-13]
43. National Health Survey: First Results, 2017-18. Australian Bureau of Statistics. 2019. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey-first-results/latest-release [accessed 2019-09-13]
44. National Aboriginal and Torres Strait Islander Social Survey, 2014-15. Australian Bureau of Statistics. 2017. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/4714.0 [accessed 2019-09-13]
45. National Health and Medical Research Council. Ethical Conduct in Research With Aboriginal and Torres Strait Islander Peoples and Communities: Guidelines for Researchers and Stakeholders. Canberra, Australia: Commonwealth of Australia; 2018.
46. Updated: Lifespan of a Social Media Post. M to M. 2019. URL: https://mtomconsulting.com/updated-lifespan-of-a-social-media-post/ [accessed 2019-10-15]
47. Freeman B, Kelly B, Baur L, Chapman K, Chapman S, Gill T, et al. Digital junk: food and beverage marketing on Facebook. Am J Public Health 2014 Dec;104(12):e56-e64. [doi: 10.2105/AJPH.2014.302167] [Medline: 25322294]
48. Hutchinson MK, Holtman MC. Analysis of count data using poisson regression. Res Nurs Health 2005 Oct;28(5):408-418. [doi: 10.1002/nur.20093] [Medline: 16163676]
49. Kite J, McGill B, Freeman B, Vineburg J, Li V, Berton N, et al. User perceptions of the campaign Facebook page: a mixed methods study. Soc Media Soc 2018 Aug 16;4(3):205630511879463. [doi: 10.1177/2056305118794639]
50. Hefler M, Kerrigan V, Freeman B, Boot GR, Thomas DP. Using Facebook to reduce smoking among Australian Aboriginal and Torres Strait Islander people: a participatory grounded action study. BMC Public Health 2019 May 21;19(1):615 [FREE Full text] [doi: 10.1186/s12889-019-6918-7] [Medline: 31113480]
51. Fogarty W, Lovell M, Langenberg J, Heron MJ. Deficit Discourse and Strengths-based Approaches: Changing the Narrative of Aboriginal and Torres Strait Islander Health and Wellbeing. Melbourne: Lowitja Institute; 2018.
52. Jeong M, Tan AS, Brennan E, Gibson L, Hornik RC. Talking about quitting: interpersonal communication as a mediator of campaign effects on smokers' quit behaviors. J Health Commun 2015;20(10):1196-1205 [FREE Full text] [doi: 10.1080/10810730.2015.1018620] [Medline: 26147367]
53. Dunlop S, Cotter T, Perez D. When your smoking is not just about you: antismoking advertising, interpersonal pressure, and quitting outcomes. J Health Commun 2014;19(1):41-56. [doi: 10.1080/10810730.2013.798375] [Medline: 23967804]
54. Campbell M, Hunt J, Scrimgeour DJ, Davey M, Jones V. Contribution of aboriginal community-controlled health services to improving aboriginal health: an evidence review. Aust. Health Review 2018;42(2):218-226. [doi: 10.1071/ah16149]
55. Thompson SC, Haynes E, Shahid S, Woods JA, Teng TH, Davidson PM, et al. Shedding light or fanning flames?: a consideration of the challenges in exploring the relative effectiveness of Aboriginal Community Controlled Health Services. Qual Primary Care 2015;23(3):141-149.

Abbreviations

ACCHO: Aboriginal Community Controlled Health Service
IRR: incident rate ratio

Edited by G Eysenbach; submitted 05.11.19; peer-reviewed by S Beasley, A Wright; comments to author 17.04.20; revised version received 23.07.20; accepted 11.11.20; published 10.12.20.

Please cite as:
Hefler M, Kerrigan V, Grunseit A, Freeman B, Kite J, Thomas DP
Facebook-Based Social Marketing to Reduce Smoking in Australia’s First Nations Communities: An Analysis of Reach, Shares, and Likes
J Med Internet Res 2020;22(12):e16927
URL: http://www.jmir.org/2020/12/e16927/
doi:10.2196/16927
PMID:33300883

Original Paper

Dimensions of Misinformation About the HPV Vaccine on Instagram: Content and Network Analysis of Social Media Characteristics

Philip M Massey[1], MPH, PhD; Matthew D Kearney[1], MPH; Michael K Hauer[1], MPH; Preethi Selvan[2], MPH; Emmanuel Koku[3], PhD; Amy E Leader[2], MPH, DrPH

[1]Department of Community Health and Prevention, Dornsife School of Public Health, Drexel University, Philadelphia, PA, United States
[2]Division of Population Science, Department of Medical Oncology, Thomas Jefferson University, Philadelphia, PA, United States
[3]Department of Sociology, College of Arts and Sciences, Drexel University, Philadelphia, PA, United States

Corresponding Author:
Philip M Massey, MPH, PhD
Department of Community Health and Prevention
Dornsife School of Public Health
Drexel University
3215 Market St
Philadelphia, PA, 19104
United States
Phone: 1 267 359 6067
Email: pmm85@drexel.edu

Abstract

Background: The human papillomavirus (HPV) vaccine is a major advancement in cancer prevention and this primary prevention tool has the potential to reduce and eliminate HPV-associated cancers; however, the safety and efficacy of vaccines in general and the HPV vaccine specifically have come under attack, particularly through the spread of misinformation on social media. The popular social media platform Instagram represents a significant source of exposure to health (mis)information; 1 in 3 US adults use Instagram.

Objective: The objective of this analysis was to characterize pro- and anti-HPV vaccine networks on Instagram, and to describe misinformation within the anti-HPV vaccine network.

Methods: From April 2018 to December 2018, we collected publicly available English-language Instagram posts containing hashtags #HPV, #HPVVaccine, or #Gardasil using Netlytic software (n=16,607). We randomly selected 10% of the sample and content analyzed relevant posts (n=580) for text, image, and social media features as well as holistic attributes (eg, sentiments, personal stories). Among antivaccine posts, we organized elements of misinformation within four broad dimensions: 1) misinformation theoretical domains, 2) vaccine debate topics, 3) evidence base, and 4) health beliefs. We conducted univariate, bivariate, and network analyses on the subsample of posts to quantify the role and position of individual posts in the network.

Results: Compared to provaccine posts (324/580, 55.9%), antivaccine posts (256/580, 44.1%) were more likely to originate from individuals (64.1% antivaccine vs 25.0% provaccine; *P*<.001) and include personal narratives (37.1% vs 25.6%; *P*=.003). In the antivaccine network, core misinformation characteristics included mentioning #Gardasil, purporting to reveal a lie (ie, concealment), conspiracy theories, unsubstantiated claims, and risk of vaccine injury. Information/resource posts clustered around misinformation domains including falsification, nanopublications, and vaccine-preventable disease, whereas personal narrative posts clustered around different domains of misinformation, including concealment, injury, and conspiracy theories. The most liked post (6634 likes) in our full subsample was a positive personal narrative post, created by a non-health individual; the most liked post (5604 likes) in our antivaccine subsample was an informational post created by a health individual.

Conclusions: Identifying characteristics of misinformation related to HPV vaccine on social media will inform targeted interventions (eg, network opinion leaders) and help sow corrective information and stories tailored to different falsehoods.

(J Med Internet Res 2020;22(12):e21451) doi:10.2196/21451

KEYWORDS

social media; cancer; vaccination; health communication; public health; HPV, human papillomavirus

Introduction

Approximately 80 million people in the United States, or about 1 in 4, are infected with human papillomavirus (HPV), and 14 million new cases of HPV will occur each year [1]. Certain strains of HPV are responsible for over 90% of anal and cervical cancers, 70% of vaginal and vulvar cancers, and 60% of penile and oropharynx cancers [1]. The HPV vaccine is a major advancement in cancer prevention, and this primary prevention tool has the potential to reduce and eliminate HPV-associated cancers.

In 2016, the National HPV Vaccination Roundtable identified social media as the top priority to strengthen vaccine confidence and increase HPV vaccination rates [2], and, more recently, the National Institutes of Health underscored its support for digital health behavior research [3]. Research on the HPV vaccine and social media has emerged across various platforms including Twitter [4-7], YouTube [8], Facebook [9], Instagram [10,11], and online media more broadly [12]. Studies show that provaccine content on social media is sizeable [5,6]; however, exposure to negative vaccine content may lead to the formation of negative opinions and subsequent sharing of antivaccine content [4], thus contributing to and perpetuating antivaccine content on social media. The HPV vaccine has come under attack, particularly through the spread of misinformation, as falsehoods and unsubstantiated claims attempt to discredit the science behind the safety and efficacy of the vaccine [13]. Social media is an important resource for health information, and, at the same time, represents a significant source of exposure to health misinformation [14].

The popular social media platform Instagram represents a significant source of exposure to health information and misinformation [14]; 1 in 3 US adults use Instagram [15]. Nearly 50% of adults ages 30-49 years use Instagram [16], corresponding to a key demographic of parents who likely have age-eligible children (9-14 years old) for the HPV vaccine and may be looking for information on social media. However, research examining Instagram data has been slower to emerge compared with other platforms, in part due to limited data access and the platform culture of posting only to closed networks of friends. To date, health-related research on Instagram has focused on characterizing images for content and theme [17] and geo-spatial analysis [18]. To our knowledge, only 2 studies have examined the HPV vaccine on Instagram [10,11], and both characterized pro- and antivaccine content. Our study extends this work by characterizing domains of misinformation among anti-HPV vaccine posts on Instagram and also by conducting a network analysis of post characteristics based on image and text features. While social network analysis has been used to better understand interactions on other social media platforms such as Facebook [19] and Twitter [20], few have studies used it to understand Instagram [21].

Misinformation, specifically spread through social media and other online platforms, is a major threat to public health and medicine [13]. Addressing misinformation on social media requires a proactive approach. To be proactive, we must first understand what types of misinformation are present. However, developing strategies to address misinformation requires more effort, including cultivating an understanding of the following: what types of messages are being shared and by whom; how the relationship between posts, hashtags, and various text/image characteristics reflect pro- and anti-HPV vaccine domains; and how to leverage these network relationships to address misinformation. To address these goals, this study uses mixed methods (qualitative and quantitative social network analysis) to examine four research questions:

1. How do Instagram post characteristics (such as format, source, and content) differ by HPV vaccine sentiment?
2. What are the salient dimensions of misinformation among anti-HPV vaccine Instagram posts?
3. What is the network structure of pro- and anti-HPV vaccine Instagram post characteristics?
4. How do position (centrality) and popularity (number of likes) of posts vary by post characteristics and domains of misinformation?

Methods

Study Design and Sampling

Between April 2018 and December of 2018, we used Netlytic [22] software to collect public Instagram posts. We collected data by accessing Instagram's public application programming interface, meeting the company's terms of service for public data, and collected up to 100 new posts per hour. If more than 100 posts with a particular hashtag/keyword were posted per hour, only the most recent were retrieved. In December 2018, Instagram closed its application programming interface and data collection through Netlytic was no longer possible [23]. As such, all data for this study were collected prior to the application programming interface closure. The data collection method in this study is similar to that used in prior Instagram research [10,18,24].

Drawing from prior social media studies on this topic [4-6] and working to maximize the number of relevant of posts (ie, signal) while limiting irrelevant posts (ie, noise), we used the hashtag search criteria "#HPV," "#HPVVaccine," and "#Gardasil". These 3 hashtags created 3 separate datasets totaling 126,327 posts. We created a merged dataset (n=48,921) after removing duplicate posts. In this merged dataset, we excluded non-English-language posts (two-thirds of the sample) using Google's translation application programming interface [25] to create a final sample of 16,607 posts. No private Instagram posts were included in our sample. All study procedures were approved by the institutional review board at Drexel University.

Content Analysis

We randomly selected 1660 of the 16,607 posts (approximately 10% of the final sample) to create a subsample for content analysis. Posts in the subsample that were not relevant to the

HPV vaccine (eg, about HPV more generally or another vaccine) were not analyzed (n=757). Additionally, posts in the subsample were not analyzed if we could not access the image through the hyperlink in our dataset (n=298). Working hyperlinks were unavailable if the post had been deleted since data collection, or if the user had changed the account privacy settings. Our content analyzed subsample included 605 posts. Manifest characteristics of posts' imagery, caption texts, and holistic post attributes (ie, source, context/style, and sentiment) were coded using a modified version of a codebook (see Multimedia Appendix 1) previously tested for reliability in analyzing HPV-related Instagram posts [10].

We organized elements of misinformation within four broad dimensions based on a review of the literature: 1) misinformation theoretical domains, 2) vaccine debate topics, 3) evidence base, and 4) health beliefs. Misinformation theoretical domains drew from Information Manipulation Theory as adapted and defined by Zhou and Zhang [26], and included concealment (ie, purporting to reveal a lie), ambivalence (ie, raising questions), distortion (ie, misrepresenting original information), and falsification (ie, fabricating information). Vaccine debate topics included common themes and ideas that are shared in antivaccine communities, including vaccine inefficacy, civil liberties, alternative medicine, ideology, and conspiracy theories [27]. Evidence base was defined as the type of information cited as the basis for assertions about the HPV vaccine, including nanopublications (eg, academic manuscripts), vaccine injury stories, and unsubstantiated claims (ie, no scientific evidence provided). Finally, we included constructs from the Health Belief Model [28] that captured risk (ie, severity and susceptibility) of vaccine-related injury and vaccine-preventable diseases, barriers and benefits of not vaccinating, and self-efficacy to not vaccinate (ie, cues to action, perceived behavioral control). All misinformation elements were coded independently and were not mutually exclusive.

Content analysis of the subsample was completed by four members of the study team who had previously analyzed HPV vaccine posts on Instagram [10]. The study team classified multiple samples of posts and resolved coding discrepancies through iterative review and refinement of the codebook. This coding method has been used widely in social media content analysis [4,5]. Our full codebook is provided as supplemental material, defining each subdomain (Multimedia Appendix 1).

To identify the presence of keywords in the caption text of posts, we searched for characters in a string that represented topics relevant to HPV vaccine characteristics (eg, "cancer," "CDC"). We also searched post metadata to identify hashtags; mentions; reposts; and, if a location was included, relevant social media characteristics. The caption text and social media characteristics were then examined vis-à-vis vaccine sentiment.

Network Analysis of Posts and Hashtags

We constructed a 2-mode affiliation network of the relationship between Instagram posts and coded HPV vaccine terms and characteristics. We constructed two networks: (a) 580 x 14 "general" network comprising 580 pro- and antivaccine posts and the 14 terms mentioned, and (b) a 256 x 23 subnetwork of 256 antivaccine posts and the 23 misinformation dimensions/domains/themes and hashtags. The cells *Xij* in each of the networks' rectangular matrices take the value "1" if a post mentions a specific term (or in the case of the antivaccine network, if a post is associated with a specific domain/dimension/theme), and "0", if otherwise. We also captured pertinent post characteristics ("attributes"); these include the presence of text/images in posts, social media features such as links, and other holistic features such as post sentiment, source, and context (for both the "general" and antivaccine networks). The resultant visualization of these networks, produced by UCINET/Netdraw [26] software's graph theoretic spring-embedding algorithm, are shown in Figures 1 and 2.

We used UCINET software [29] to compute degree centrality, an indicator of how connected or popular a single node is and how likely such a node is in transmission of information through a network. Degree centrality measures the absolute number of other nodes that each node is connected to. Additionally, UCINET's core-periphery procedures determined the presence of key content clusters, distinguishing between "core" and "peripheral" characteristics in the general- and antivaccine networks. Core-periphery analysis allows for the examination of the extent to which groups are clustered and communicate about issues of mutual interest, as well as how content clusters are grouped around diverse and loosely connected sets of topics, posts, or issues [30].

Figure 1. Two-mode visualization (n=580 posts; neutral posts excluded). Includes image, caption, and social media characteristics. Variables colored by type of characteristic. Sized by likes (mean=145.8; median=21; maximum=6634). Top two posts with the most likes are indicated. Symbol shapes represent post source. Color represents node type. Rim color indicates post context. Yellow = social media features. Light blue = image characteristics. Dark blue = caption text characteristics. Red = antivaccine. Green = provaccine. Black rim = personal narrative. White rim = information/resource. Circle = general group. Square = general individual. Triangle = health group. Diamond = health individual.

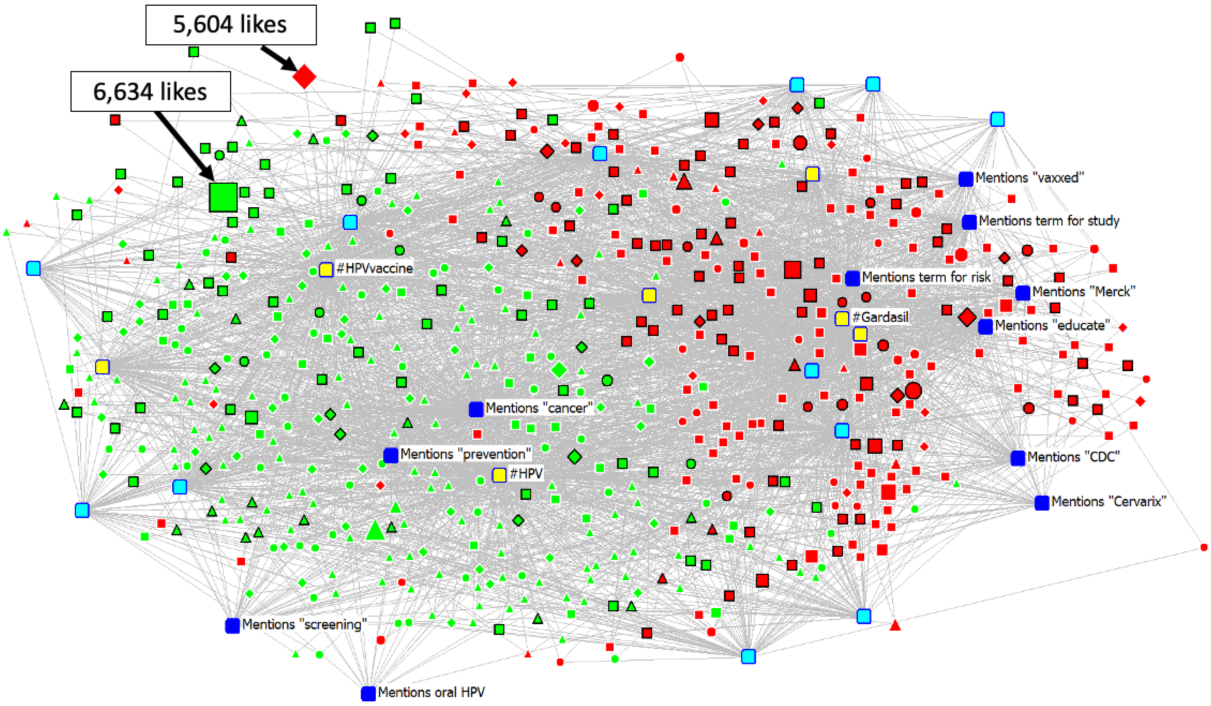
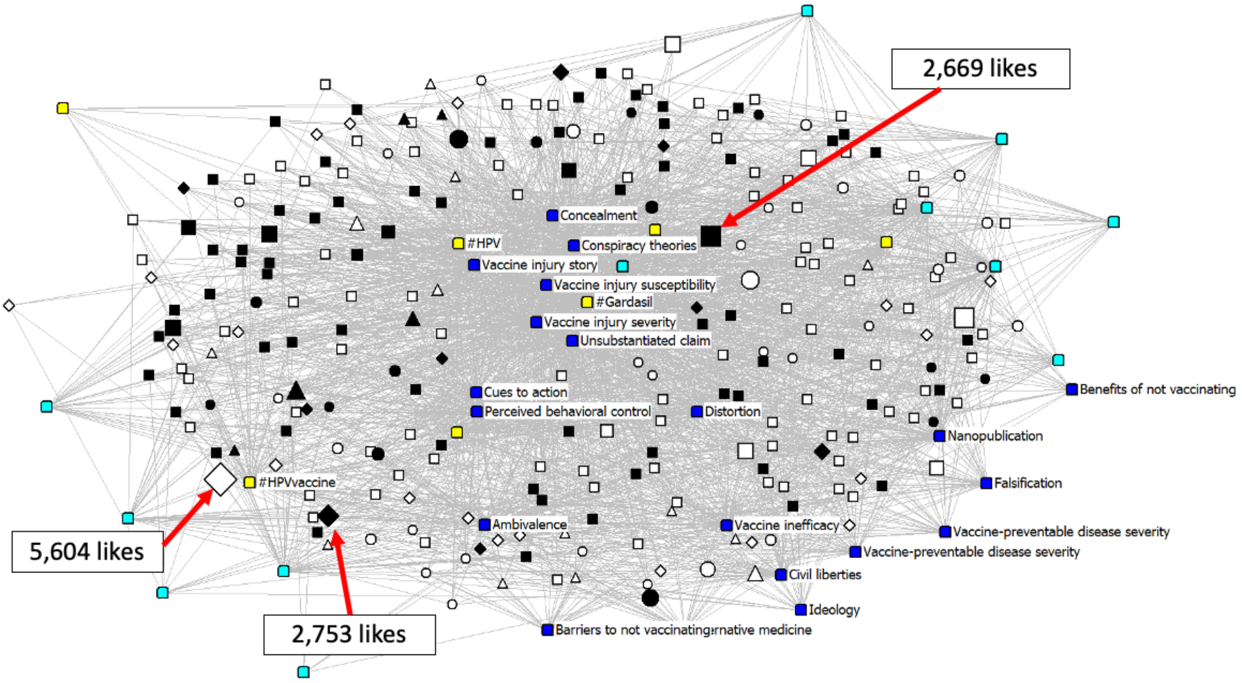


Figure 2. Antivaccine network visualization (n=256 posts). Variables colored by type of characteristic. Sized by likes (mean=220.9; median=27; maximum=5604). Top three posts with the most likes are indicated. Includes image, type of misinformation, and social media characteristics. Symbol shapes represent post source. Color represents node type. Yellow = social media features. Light blue = image characteristics. Dark blue = type of misinformation. Black = personal narrative. White = information/resource. Circle = general group. Square = general individual. Triangle = health group. Diamond = health individual.



Quantitative Analysis

Statistical analyses included both univariate and bivariate analyses and were conducted in STATA/IC Version 15 software [31]. Simple descriptive statistics were generated for pro- and antivaccine posts separately and in aggregate. Among both pro- and antivaccine posts, *t* tests and analysis of variance (ANOVA) assessed differences in 'like' count by post characteristics. chi-squared tests assessed differences in the distribution of characteristics between pro- and antivaccine posts.

Results

Inclusion and Exclusion

Of the relevant and working 605 posts in our subsample, a small proportion (n=25, 4.1% coded posts) were determined to be neutral (ie, neither pro- nor antivaccine), and were therefore excluded from subsequent analyses. Thus, the final analytic sample included 256 antivaccine posts and 324 provaccine posts (n=580 total).

Content Analysis

Intercoder reliability was assessed using percentage agreement throughout the codebook development process (mean agreement=0.87, SD=0.09). Following codebook development, members of the study were randomly assigned unique samples of posts to code, as well as an additional sample of shared posts to again evaluate intercoder reliability (mean agreement=0.85, SD=0.09), ensuring our coding process was rigorous and replicable. Agreement for individual codes ranged from 0.67 (susceptibility of vaccine-related injury) to 1.0 (barriers to not vaccinating).

Table 1 summarizes coded post characteristics and social media features, stratified by vaccine sentiment. The majority of posts were provaccine (324/580, 55.9%). Individuals were the most frequent type of post source (245/580, 42.2%), followed by health non-individuals (151/580, 26.0%, general non-individuals (102/580, 17.6%), and health individuals (82/580, 14.1%). Nearly 7 in 10 posts contained information or resources (402/580, 69.3%) compared to personal narratives (178/580, 30.7%). Many post visuals were either imagery-only (187/580, 32.2%) or noninfographic combinations of text and imagery (173/580, 29.8%). Most posts included at least one person (329/580, 56.7%), such as a vaccine-eligible child (269/580, 46.4%), a health professional (91/580, 15.7%), or parent/caregiver (30/580, 5.2%). Females were depicted more than twice as frequently as males (females=134/580, 23.1%; males=58/580, 10.0%). Antivaccine posts received significantly more likes compared to provaccine posts (220.9 mean likes vs 86.3 mean likes, respectively; *P*=.003). Posts containing personal narratives received significantly more likes compared to posts containing information/resources (217.5 mean likes vs 114.0 mean likes, respectively; *P*=.033).

Significant differences were found between characteristics of pro- and antivaccine posts (see Table 1). Compared to provaccine posts, antivaccine posts were more likely to originate from non-health individuals (164/256, 64.1% antivaccine vs 81/324, 25.0% provaccine; *P*<.001), include personal narrative (95/256, 37.1% vs 83/324, 25.6%; *P*=.003), or show a parent/caregiver (21/256, 8.2% vs 9/324, 2.8%; *P*=.003). Antivaccine posts were also more likely to mention another Instagram user (115/256, 44.9% antivaccine vs 87/324, 26.9% provaccine; *P*<.001), include a link (75/256, 29.3% vs 51/324, 15.7%; *P*<.001), or be a repost of another post (79/256, 30.9% vs 23/324, 7.1%; *P*<.001). Provaccine posts were more likely than antivaccine posts to include location information (80/324, 24.7% vs 9/256, 3.5%; *P*<.001). Finally, vaccine sentiment was a significant determinant of which hashtags were included, with antivaccine posts using #Gardasil significantly more often than provaccine posts (*P*<.001), and provaccine posts using #HPV (*P*<.001) and #HPVvaccine (*P*<.001) more than antivaccine posts.

Table 2 presents coding results and mean like counts for misinformation elements from antivaccine posts only (n=256). Concealment and distortion were the most frequent misinformation theoretical domains (135/256, 52.7% and 84/256, 32.8%, respectively). The most common vaccine debate topics were conspiracy theories (144/256, 56.3%) and vaccine inefficacy (72/256, 28.1%). Nearly three quarters of antivaccine posts offered unsubstantiated claims (185/256, 72.3%). The majority of posts highlighted the risk of vaccine-related injury: approximately 8 in 10 (205/256, 80.1%) discussed severity and approximately 6 in 10 (163/256, 63.7%) discussed susceptibility. One fifth of posts discussed the risks of vaccine-preventable diseases, primarily by downplaying susceptibility (41/256, 16.0%) of vaccine-preventable diseases. Building self-efficacy to not vaccinate was another key component of posts: 40.2% (103/256) of posts promoted one's behavioral control over not vaccinating, and 39.8% (102/256) mentioned tangible cues to action such as links to vaccine exemption forms.

Table 1. Frequency and proportion of Instagram post characteristics (ie, image, text, and social media features) for pro- and antivaccine posts. Results ordered descending by total frequency. Mean like counts and SDs are provided for all posts. Chi-squared tests assessed significant differences in distribution of post characteristics between pro- and antivaccine posts (alpha=.05).

Post Characteristics	Value (N=580), n (%)	Mean likes, n (SD)	Provaccine (n=324), %	Antivaccine (n=256), %	*P* value
Whole-post attributes					
Vaccine sentiment			**N/A**[a]	**N/A**	**N/A**
Provaccine	324 (55.9)	86.3 (484.6)			
Antivaccine	256 (44.1)	220.9 (591.0)			
Neutral[b]	25 (4.1)	29.4 (67.8)			
Post source					**<.001**
General individual	245 (42.2)	132.2 (539.8)	25	64.1	
Health non-individual	151 (26.0)	109.2 (464.8)	41.1	7	
General non-individual	102 (17.6)	174.8 (418.5)	17.6	17.6	
Health individual	82 (14.1)	217.5 (749.9)	16.4	11.3	
Post content					**0.003**
Personal narrative	178 (30.7)	217.5 (673.7)	25.6	37.1	
Information/resource	402 (69.3)	114 (462.7)	74.4	62.9	
Image characteristics					
Visualization					**<.001**
Imagery-only	187 (32.2)	135.4 (578.4)	43.2	18.4	
Text + imagery: noninfographic	173 (29.8)	181.8 (459.2)	18.5	44.1	
Text-only	74 (12.8)	207.9 (891.6)	13	12.5	
Text + imagery: infographic	62 (10.7)	75.7 (268.6)	17.3	2.3	
Video	58 (10.0)	93.7 (296.3)	7.1	13.7	
Other	26 (4.5)	86.8 (141.0)	0.9	9	
Total person(s) shown					**0.393**
None	251 (43.3)	136.1 (558.5)	46.3	39.5	
1 person	199 (34.3)	184.5 (615.4)	31.8	37.5	
2-9 persons	117 (20.2)	100.7 (330.4)	19.8	20.7	
10+ persons	13 (2.2)	143.3 (332.0)	2.2	2.3	
Vaccine-eligible person shown					**0.02**
None	311 (53.6)	146.4 (545.4)	54.9	52	
Female(s) only	134 (23.1)	166.2 (671.1)	20.4	26.6	
Male(s) only	58 (10.0)	172.2 (398.8)	9	11.3	
Both male(s) and female(s)	47 (8.1)	108.8 (329.9)	11.1	4.3	
Unable to determine	30 (5.2)	54.8 (177.3)	4.6	5.9	
Other image elements					
Vaccine shown	146 (25.2)	158.7 (472.0)	22.8	28.1	0.15
Health professional shown	91 (15.7)	110.4 (347.7)	19.4	10.9	0.005
Parent/caregiver shown	30 (5.2)	177.3 (489.6)	2.8	8.2	0.003
Hashtag(s) mentioned					
#HPV	330 (56.9)	97.6 (374.9)	74.1	35.2	<.001
#HPVVaccine	271 (46.7)	133.1 (643.3)	60.5	29.3	<.001
#Gardasil	268 (46.2)	179.2 (475.9)	23.2	75.4	<.001

Post Characteristics	Value (N=580), n (%)	Mean likes, n (SD)	Provaccine (n=324), %	Antivaccine (n=256), %	*P* value
#HPV + #HPVVaccine	121 (20.9)	90.1 (450.9)	35.19	2.73	<.001
#HPV + #Gardasil	96 (16.6)	148.4 (379.6)	11.7	22.7	<.001
#HPVVaccine + #Gardasil	37 (6.4)	97.1 (368.3)	1.8	12.1	<.001
#HPV + #HPVVaccine + #Gardasil	26 (4.5)	151.1 (486.3)	5.6	3.1	0.16
Social media characteristics					
Other user mentioned	202 (34.8)	123.5 (362.3)	26.9	44.9	<.001
Link included	126 (21.7)	142.4 (446.5)	15.7	29.3	<.001
Post is a repost	102 (17.6)	110 (336)	7.1	30.9	<.001
Location included	89 (15.3)	78.6 (312.1)	24.7	3.5	<.001

[a] N/A: Not applicable.

[b] Neutral posts excluded from subsequent analyses.

Table 2. Frequency and proportion of misinformation characteristics of antivaccine Instagram posts (n=256). Results ordered descending by frequency. All characteristics were coded independently and were not mutually exclusive. Not shown (<10%): severity of vaccine-preventable diseases, benefits of not vaccinating, barriers to not vaccinating.

Category, characteristic	Value, n	Mean likes, n (SD)	Degree centrality[a]
Misinformation domains			
Concealment	135	238.3 (677.7)	0.527
Distortion	84	167.4 (391.1)	0.328
Ambivalence	73	155.1 (391.3)	0.285
Falsification	40	267.6 (525.0)	0.156
Vaccine debate topics			
Conspiracy theories	144	152.8 (378.7)	0.563
Vaccine inefficacy	72	254.3 (632.0)	0.281
Civil liberties	49	193 (491.5)	0.191
Alternative medicine	34	257 (573.8)	0.133
Ideological	26	243 (548.2)	0.102
Evidence base			
Unsubstantiated claim	185	156.8 (416.1)	0.723
Vaccine-injury stories	116	209.4 (482.4)	0.453
Nanopublication	71	254.5 (789.6)	0.277
Health beliefs			
Severity of vaccine-related injury	205	214.9 (599.8)	0.801
Susceptibility of vaccine-related injury	163	192.5 (464.8)	0.637
Perceived behavioral control	103	306.8 (772.9)	0.402
Cues to action	102	286.9 (771.1)	0.398
Susceptibility to vaccine-preventable diseases	41	253 (594.2)	0.16

[a] Degree centrality an indicator of how connected a single characteristic is and how likely such a characteristic is in transmission of information and resources through a network; the higher the measure the more common or frequently occurring the feature is in the network. Average degree centrality in antivaccine network was 0.336 (SD=0.219).

Network Analysis

Figure 1 presents the 2-mode network visualization of the coded subsample (n=580 posts). Different hashtags were used preferentially depending on a post's vaccine sentiment. #Gardasil was core to the antivaccine network, whereas #HPVvaccine and #HPV were both core to the provaccine network. Text including risk, study, and educate all gravitated towards #Gardasil and were central to the antivaccine network, whereas text including prevention and cancer both gravitated towards #HPV and were central to the provaccine network. Although provaccine posts were more common, antivaccine posts included many more nodes with black rims, indicating that the post was personal narrative as opposed to information/resource (displayed with a white rim). The post nodes were sized proportionate to the number of likes received – the most liked post (n=6634 likes) in our subsample was a positive personal narrative post, created by a non-health individual (located in the upper left quadrant of the network diagram).

Figure 2 presents only the antivaccine network (n=256 posts) with additional characteristics of misinformation. The position of theoretical dimensions of misinformation varied. Concealment was located in the core of the network, whereas distortion, falsification, and ambivalence were more peripheral. In addition, unsubstantiated claims and vaccine injury stories were core pieces of evidence to the network. With respect to health belief model constructs, risk of vaccine injury and self-efficacy to not vaccinate were located in the core of the antivaccine network. Finally, posts that tap on conspiracy theories and vaccine inefficacy were central and located in the core of the network, whereas those tapping on vaccine debate topics, ideology, alternative medicine, and civil liberties were in the periphery. As shown in Figure 2, social media features (colored in yellow) such as use of hashtags, links, and mentioning other users were also located in the center of the network. Information/resource posts (colored in white) clustered around misinformation domains including falsification, nanopublications, and vaccine-preventable disease; whereas personal narrative posts (colored in black) clustered around different domains of misinformation, including concealment (ie, revealing lies), injury, and conspiracy theories. Finally, post nodes were sized proportionate to the number of likes received – the most liked post (n=5604 likes) in our antivaccine subsample was an informational post created by a health individual (a white diamond located in the lower left quadrant of the network diagram).

On average, degree centrality in the antivaccine network was 0.336 (SD=0.219). As shown in Table 2, degree centrality for posts ranged from 0.102 (ideological) to 0.801 (vaccine injury severity). Common misinformation elements of the antivaccine network with degree centrality scores greater than 0.5 included concealment (0.527), conspiracy theories (0.563), unsubstantiated claim (0.723), severity of vaccine-related injury (0.801), and susceptibility of vaccine-related injury (0.637). Instagram posts that exemplify the common misinformation elements are included as Multimedia Appendix 2.

Discussion

Instagram Post Characteristics by HPV Vaccine Sentiment

The majority of Instagram posts in our HPV vaccine sample were provaccine and used hashtags #HPV and #HPVvaccine. Antivaccine posts received on average more likes than provaccine and were more likely to use #Gardasil. Use of social media features also varied by post sentiment. Antivaccine posts were more likely to mention another Instagram user (ie, direct communication), and provaccine posts were more likely to include location information – suggesting differences in how the two groups connect with others and share information. For example, antivaccine posts included location information significantly less often than provaccine posts, and the presence or absence of geotagging may be an important marker for the transparency and credibility of content creators [32]. Finally, in our sample, more posts contained information/resources compared to personal narratives, and the latter received more likes on average, demonstrating the power and popularity of a story. These results confirm and extend findings from a study that used a more limited dataset [10].

Our findings examining Instagram data support research conducted on other social media platforms related to the HPV vaccine. Similar to our findings, Twitter studies have found the majority of content to be provaccine [5,6]. On YouTube, pro-HPV vaccine content relied heavily on information and evidence (as compared to personal stories), and antivaccine content on YouTube focused on side effects and conspiracy theories [8]. While our findings support prior work on other social media platforms, it also extends this knowledge base by examining misinformation domains.

Dimensions of Misinformation Among Anti-HPV Vaccine Posts

Our network diagram related to misinformation among antivaccine posts (Figure 2) not only highlights domains and topics that cluster together (eg, conspiracy, injury, and concealment vs behavioral control, vaccine inefficacy, and distortion) but also the post characteristics that group around each cluster (eg, personal narrative vs information/resources). This may inform public health messaging to better pair with existing content. For instance, personal narrative posts may be better suited to address conspiracy than information/resource posts. In our antivaccine sample, misinformation was represented through core elements, including posts using concealment strategies, posts highlighting conspiracy theories, posts basing claims on unsubstantiated evidence and personal anecdotes (eg, injury), and posts raising awareness of vaccine-related injury severity and susceptibility. Multimedia Appendix 2 provides three sample posts that exemplify each of the core misinformation elements.

Concealment was the core misinformation theoretical domain in our subsample's network. Posts that used concealment as the vehicle for misinformation purported to reveal a lie or expose previously unknown facts. An example provided from our data demonstrated that a "new study" helped to reveal previously "unknown facts" about the HPV vaccine trials. While not a core component of the network, distortion also demonstrated high degree centrality. Whereas concealment posts shined a light on information, distortion posts created false light, presenting one or more potentially true pieces of evidence to imply correlation, causation, or comparison between them. Distortion was particularly salient among injury stories that drew links between receiving a vaccine and injury. The concealment and distortion misinformation domains warrant unique strategies to address hesitancy, fear, or doubt that may arise from exposure to such information.

Conspiracy theories were also core to the antivaccine network and included posts that claimed various actors (eg, government, nonprofit, or industry) wanted to promote HPV vaccination for nefarious reasons. Conspiracy is not a new topic among vaccines or health more broadly, and strategies to address this type of misinformation may seek to identify hidden agendas or groups with self-serving interests, including but not limited to financial interests. Other strategies to address conspiracy may be to identify groups that feel alienated and create opportunities for dialogue [33].

A third core component of misinformation posts was the use of unsubstantiated or anecdotal evidence (eg, personal experiences) to corroborate falsehoods. Posts that claimed vaccines cause autism or SIDS (sudden infant death syndrome), for instance, are examples of unsubstantiated evidence as there is scientific evidence to support the contrary [34]. On the other hand, many posts focused on injury as evidence for antivaccine sentiment based on personal or anecdotal experiences. Public health professionals have better developed tools to address misinformation that is “evidence based” or falsehoods that are based on shaky science. However, the tools or resources needed to address an antivaccine story about a personal experience may need to better incorporate emotional evidence, acknowledging the struggle or emotions elicited by the story and using these same emotions to redirect the narrative. Medical professionals are taught to show empathy towards a patient; public health professionals can draw from this to build additional tools to combat misinformation on social media.

Finally, as related to health beliefs, misinformation posts focused on severity and susceptibly of vaccine-related injury. Severity posts fixated on harmful side effects, illnesses, and even death – not to mention possible unknown long-term effects. Susceptibly, on the other hand, included arguments about the commonality of side effects and the number of vaccines on the childhood schedule, therefore overloading the immune system. This last sentiment may be particularly salient to parents who are concerned with the HPV vaccine being routinely given with the meningococcal and Tdap vaccines.

Limitations

Our study has limitations worth noting. First, our sample was created using three hashtags, limiting the generalizability of our findings. We mitigated this potential bias by using three common hashtags that have been used in other studies and have been shown to include both pro- and antivaccine content [4,5]. Second, we examined misinformation domains only among antivaccine posts. While provaccine posts may contain false information, we focused on antivaccine posts. Third, our unobtrusive methodology is unable to measure actual exposure to posts or determine if exposure was associated with health knowledge, attitudes, or practices. Additionally, Instagram’s application programming interface did not generate user information beyond the user’s account name, so follower counts were not available during our analysis as a measure of potential exposure. Future research may consider examining differences in posts from users with larger versus smaller followings. Fourth, we included mean and not median like count in our analysis. The median number of likes for antivaccine posts was 27 (mean=220.9), and for provaccine posts was 18 (mean=86.3), confirming prior reporting that engagement measures such as likes are right-skewed because a minority of posts receive a disproportionate share of likes [35];this is a typical phenomenon in social media research and practice. Although the mean and median like counts differ in terms of magnitude, they do not differ directionally. Moreover, posts were created at different times and therefore had varying amounts of time to accumulate likes. However, because most social media engagement occurs within a short amount of time immediately following the creation of a post, the impact on our findings was likely minimal. Furthermore, in our analytic sample, like count was not associated with the number of posts created by a given user (P=.909; results not presented), suggesting that no *one* user dominated or influenced the number of likes. Additionally, our decision to randomly sample posts did not allow for us to examine temporal trends and future studies may consider a stratified sampling approach by week or another unit of time. Finally, we did not identify bot activity nor attempt to identify automation or nonhuman interactions. This activity could lead to artificially high engagement with specific types of content and warrants additional investigation to determine the presence and proliferation of misinformation resulting from automated activity. Despite these limitations, our findings help characterize misinformation about the HPV vaccine on Instagram and provide a footing for future research in this field.

Conclusion

Health misinformation on social media is diverse, tapping into states of reason to emotion. Identifying characteristics of health misinformation on social media will help inform targeted interventions and tailored messages to sow corrective information and stories [36]. The American Heart Association’s ReSS (Resuscitation Science Symposium) social media campaign [37] used a small group of resuscitation science professionals to create (corrective) content in online social media platforms, leading to significant end user engagement with the content. Similar interventions have been documented in other settings [38]. If the public health and medical community wants to be at the center of the social media network and discussion about the HPV vaccine, it must understand and consider similar strategies. Misinformation characteristics can be identified and segmented for focused interventions through opinion-leader or peer outreach education programs.

Communication strategies that only leverage conventional health experts and authorities are ill-equipped to address misinformation on social media. The rise of “expert patients” and “expert parents” has been in part due to their proficient use of social media network features, along with the saliency and relatedness of their stories. Addressing misinformation on social media will require resource development and enthusiasm across multiple industries and health consumer types, including tech and health insurance companies, hospital and physician groups, and parent and cancer survivor advocates.

Acknowledgments

PM was the lead author and wrote the manuscript, conceptualized the research questions, and colead all data analysis and interpretation. MK lead data collection and colead data analysis and writing the results. MH and PS assisted with data analysis and results. EK lead social network data analysis and interpretation. AL assisted with writing the manuscript, data analysis, and conceptualizing research questions. This study was supported by the National Cancer Institute (3P30CA056036-18S3 and R01-CA229324-01A).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook for content analysis.
[PDF File (Adobe PDF File), 120 KB - jmir_v22i12e21451_app1.pdf]

Multimedia Appendix 2
Exemplar anti-vaccine Instagram posts with defined misinformation elements.
[PDF File (Adobe PDF File), 1167 KB - jmir_v22i12e21451_app2.pdf]

References

1. Centers FDC. Genital HPV Infection - Fact Sheet. STD Fact Sheet Internet. 2017. URL: https://www.cdc.gov/std/hpv/stdfact-hpv.htm [accessed 2017-01-15]
2. Reiter PL, Gerend MA, Gilkey MB, Perkins RB, Saslow D, Stokley S, et al. Advancing Human Papillomavirus Vaccine Delivery: 12 Priority Research Gaps. Acad Pediatr 2018 Mar;18(2S):S14-S16 [FREE Full text] [doi: 10.1016/j.acap.2017.04.023] [Medline: 29502629]
3. Riley WT, Oh A, Aklin WM, Wolff-Hughes DL. National Institutes of Health Support of Digital Health Behavior Research. Health Educ Behav 2019 Dec 19;46(2_suppl):12-19. [doi: 10.1177/1090198119866644] [Medline: 31742453]
4. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]
5. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying Multiple Data Collection Tools to Quantify Human Papillomavirus Vaccine Communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
6. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. J Med Internet Res 2016 Aug 29;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
7. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017 May 25;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
8. Ekram S, Debiec K, Pumper M, Moreno M. Content and Commentary: HPV Vaccine and YouTube. J Pediatr Adolesc Gynecol 2019 Apr;32(2):153-157. [doi: 10.1016/j.jpag.2018.11.001] [Medline: 30445163]
9. Mohanty S, Leader AE, Gibeau E, Johnson C. Using Facebook to reach adolescents for human papillomavirus (HPV) vaccination. Vaccine 2018 Sep 25;36(40):5955-5961. [doi: 10.1016/j.vaccine.2018.08.060] [Medline: 30172634]
10. Kearney MD, Selvan P, Hauer MK, Leader AE, Massey PM. Characterizing HPV Vaccine Sentiments and Content on Instagram. Health Educ Behav 2019 Dec 19;46(2_suppl):37-48. [doi: 10.1177/1090198119859412] [Medline: 31742459]
11. Basch CH, MacLean SA. A content analysis of HPV related posts on instagram. Hum Vaccin Immunother 2019 Jan 30;15(7-8):1476-1478 [FREE Full text] [doi: 10.1080/21645515.2018.1560774] [Medline: 30570379]
12. Margolis MA, Brewer NT, Shah PD, Calo WA, Gilkey MB. Stories about HPV vaccine in social media, traditional media, and conversations. Prev Med 2019 Jan;118:251-256. [doi: 10.1016/j.ypmed.2018.11.005] [Medline: 30414396]
13. Chou WS, Oh A, Klein WMP. Addressing Health-Related Misinformation on Social Media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
14. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
15. Social Media Fact Sheet Internet. Pew Research Center. 2019. URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-07-01]

16. Social media usage in the U.S. in 2019 Internet. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-07-01]
17. Czaplicki L, Kostygina G, Kim Y, Perks SN, Szczypka G, Emery SL, et al. Characterising JUUL-related posts on Instagram. Tob Control 2020 Nov 02;29(6):612-617. [doi: 10.1136/tobaccocontrol-2018-054824] [Medline: 31266903]
18. Allem J, Chu K, Cruz TB, Unger JB. Waterpipe Promotion and Use on Instagram: #Hookah. Nicotine Tob Res 2017 Oct 01;19(10):1248-1252 [FREE Full text] [doi: 10.1093/ntr/ntw329] [Medline: 28077449]
19. Tulin M, Pollet TV, Lehmann-Willenbrock N. Perceived group cohesion versus actual social structure: A study using social network analysis of egocentric Facebook networks. Soc Sci Res 2018 Aug;74:161-175. [doi: 10.1016/j.ssresearch.2018.04.004] [Medline: 29961483]
20. Gruzd A, Paulin D, Haythornthwaite C. Analyzing Social Media And Learning Through Content And Social Network Analysis: A Faceted Methodological Approach. Learning Analytics 2016 Dec 19;3(3):46-71. [doi: 10.18608/jla.2016.33.4]
21. Santarossa S, Woodruff SJ. #LancerHealth: Using Twitter and Instagram as a tool in a campus wide health promotion initiative. J Public Health Res 2018 Feb 05;7(1):1166 [FREE Full text] [doi: 10.4081/jphr.2018.1166] [Medline: 29780763]
22. Gruzd A. Netlytic: Software for Automated Text and Social Network Analysis. Netlytic. 2018. URL: http://netlytic.org [accessed 2018-12-01]
23. Data source: Instagram Internet. Netlytic. 2018. URL: https://netlytic.org/home/?page_id=254 [accessed 2018-12-11]
24. Allem J, Escobedo P, Chu K, Boley Cruz T, Unger JB. Images of Little Cigars and Cigarillos on Instagram Identified by the Hashtag #swisher: Thematic Analysis. J Med Internet Res 2017 Jul 14;19(7):e255 [FREE Full text] [doi: 10.2196/jmir.7634] [Medline: 28710057]
25. Ooms J. Google's Compact Language Detector 3 Internet. CLD3 C++ library. URL: https://rdrr.io/cran/cld3/ [accessed 2020-03-01]
26. Zhou L, Zhang D. An Ontology-Supported Misinformation Model: Toward a Digital Misinformation Library. IEEE Trans. Syst., Man, Cybern. A 2007 Sep;37(5):804-813. [doi: 10.1109/tsmca.2007.902648]
27. Six common misconceptions about immunization Internet. World Health Organization (WHO). 2018. URL: https://www.who.int/news-room/q-a-detail/vaccines-and-immunization-myths-and-misconceptions [accessed 2019-10-01]
28. Rosenstock IM. Historical Origins of the Health Belief Model. Health Education Monographs 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
29. Borgatti S, Everett M, Freeman L. Ucinet for Windowsoftware for Social Network. Harvard, MA: Analytic Technologies; 2002.
30. Borgatti SP, Everett MG. Models of core/periphery structures. Social Networks 2000 Oct;21(4):375-395. [doi: 10.1016/s0378-8733(99)00019-2]
31. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017. URL: https://www.stata.com/ [accessed 2019-10-01]
32. Wirtz JG, Zimbres TM. A systematic analysis of research applying 'principles of dialogic communication' to organizational websites, blogs, and social media: Implications for theory and practice. Journal of Public Relations Research 2018 Apr 19;30(1-2):5-34. [doi: 10.1080/1062726x.2018.1455146]
33. Swami V, Barron D, Weis L, Voracek M, Stieger S, Furnham A. An examination of the factorial and convergent validity of four measures of conspiracist ideation, with recommendations for researchers. PLoS One 2017 Feb 23;12(2):e0172617 [FREE Full text] [doi: 10.1371/journal.pone.0172617] [Medline: 28231266]
34. Institute of Medine. Adverse Effects of Vaccines: Evidence and Causality. Washington, DC: The National Academies Press; 2012.
35. Hong C, Chen Z, Li C. "Liking" and being "liked": How are personality traits and demographics associated with giving and receiving "likes" on Facebook? Computers in Human Behavior 2017 Mar;68:292-299. [doi: 10.1016/j.chb.2016.11.048]
36. Valente TW. Network interventions. Science 2012 Jul 06;337(6090):49-53. [doi: 10.1126/science.1217330] [Medline: 22767921]
37. Leary M, McGovern S, Dainty KN, Doshi AA, Blewer AL, Kurz MC, et al. Examining the Use of a Social Media Campaign to Increase Engagement for the American Heart Association 2017 Resuscitation Science Symposium. In: JAHA. 2018 Apr 17 Presented at: The Resuscitation Science Symposium; 2018; New Orleans, Louisiana. [doi: 10.1161/jaha.118.008710]
38. Shi J, Salmon CT. Identifying Opinion Leaders to Promote Organ Donation on Social Media: Network Study. J Med Internet Res 2018 Jan 09;20(1):e7 [FREE Full text] [doi: 10.2196/jmir.7643] [Medline: 29317384]

Abbreviations

HPV: Human papillomavirus

Edited by G Eysenbach; submitted 15.06.20; peer-reviewed by C Basch, JP Allem; comments to author 28.10.20; revised version received 03.11.20; accepted 03.11.20; published 03.12.20.

Please cite as:
Massey PM, Kearney MD, Hauer MK, Selvan P, Koku E, Leader AE
Dimensions of Misinformation About the HPV Vaccine on Instagram: Content and Network Analysis of Social Media Characteristics
J Med Internet Res 2020;22(12):e21451
URL: https://www.jmir.org/2020/12/e21451
doi:10.2196/21451
PMID:33270038

Original Paper

Evaluating Behavioral and Linguistic Changes During Drug Treatment for Depression Using Tweets in Spanish: Pairwise Comparison Study

Angela Leis[1*], PsyM; Francesco Ronzano[1*], PhD; Miguel Angel Mayer[1], MD, MPH, PhD; Laura I Furlong[1], PhD; Ferran Sanz[1], Prof Dr

Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute, Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona, Spain

[*]these authors contributed equally

Corresponding Author:
Ferran Sanz, Prof Dr
Research Programme on Biomedical Informatics
Hospital del Mar Medical Research Institute
Department of Experimental and Health Sciences, Pompeu Fabra University
Barcelona Biomedical Research Park
Carrer Dr Aiguader 88
Barcelona, 08003
Spain
Phone: 34 933160540
Fax: 34 933160550
Email: ferran.sanz@upf.edu

Abstract

Background: Depressive disorders are the most common mental illnesses, and they constitute the leading cause of disability worldwide. Selective serotonin reuptake inhibitors (SSRIs) are the most commonly prescribed drugs for the treatment of depressive disorders. Some people share information about their experiences with antidepressants on social media platforms such as Twitter. Analysis of the messages posted by Twitter users under SSRI treatment can yield useful information on how these antidepressants affect users' behavior.

Objective: This study aims to compare the behavioral and linguistic characteristics of the tweets posted while users were likely to be under SSRI treatment, in comparison to the tweets posted by the same users when they were less likely to be taking this medication.

Methods: In the first step, the timelines of Twitter users mentioning SSRI antidepressants in their tweets were selected using a list of 128 generic and brand names of SSRIs. In the second step, two datasets of tweets were created, the *in-treatment* dataset (made up of the tweets posted throughout the 30 days after mentioning an SSRI) and the *unknown-treatment* dataset (made up of tweets posted more than 90 days before or more than 90 days after any tweet mentioning an SSRI). For each user, the changes in behavioral and linguistic features between the tweets classified in these two datasets were analyzed. 186 users and their timelines with 668,842 tweets were finally included in the study.

Results: The number of tweets generated per day by the users when they were in treatment was higher than it was when they were in the *unknown-treatment* period (P=.001). When the users were in treatment, the mean percentage of tweets posted during the daytime (from 8 AM to midnight) increased in comparison to the *unknown-treatment* period (P=.002). The number of characters and words per tweet was higher when the users were in treatment (P=.03 and P=.02, respectively). Regarding linguistic features, the percentage of pronouns that were first-person singular was higher when users were in treatment (P=.008).

Conclusions: Behavioral and linguistic changes have been detected when users with depression are taking antidepressant medication. These features can provide interesting insights for monitoring the evolution of this disease, as well as offering additional information related to treatment adherence. This information may be especially useful in patients who are receiving long-term treatments such as people suffering from depression.

(J Med Internet Res 2020;22(12):e20920) doi:10.2196/20920

KEYWORDS

depression; antidepressant drugs; serotonin uptake inhibitors; mental health; social media; infodemiology; data mining

Introduction

Background

Depression is one of the most common mental disorders [1]. According to the World Health Organization, depression affects more than 322 million people of all ages globally, being a leading cause of disability worldwide [2]. The proportion of people with depression went up by around 18% between 2005 and 2015 [3]. This mental disorder constitutes a challenge for society and health care systems due to devastating personal and social consequences and the associated economic costs [4-13]. In spite of the high prevalence of depression and the efforts of health care services to improve its management, this health condition remains underdiagnosed and undertreated [14].

In the case of moderate and severe forms of depression, pharmacological treatment can improve the quality of life of these patients [4]. There are several types of antidepressant drugs, and among them, selective serotonin reuptake inhibitors (SSRIs) are currently the most prescribed antidepressants around the world. For instance, according to the Spanish Agency for Medicines and Health Products [15], SSRIs constitute more than 70% of all antidepressants prescribed in Spain. They have fewer side effects than other antidepressants [16], show a good risk-benefit ratio [17,18], are safer and better tolerated [19], and exhibit a reduced risk of toxicity in overdose in comparison to tricyclic antidepressants [20]. They are commonly used as first-line treatment for depression [21-23] and are usually prescribed as maintenance therapy to prevent relapse [4,23-26]. SSRIs include the following drugs: fluvoxamine, fluoxetine, paroxetine, sertraline, citalopram, and escitalopram [17].

Furthermore, although social media platforms have typically not been created with health-related purposes in mind [27,28], millions of people publicly share personal health information on social media platforms every day [29,30]. For this reason, these platforms represent an important source of health information that is faster and more broadly available than other sources of health information, being unsolicited, spontaneous, and up to date. Infodemiology approaches have been developed and applied to better understand the dynamics of these platforms when used as a health information source [31-33]. In this context, social media users share health-related information, such as experiences with prescribed drugs [34], cancer patients' sentiments [35], opinions on vaccines [36], or online conversations on epidemic outbreaks [37]. The massive data from social media can be monitored and analyzed by using natural language processing and machine learning technologies, providing new possibilities to better understand users' behavior [30], including automatic identification of early signs of mental disorders [38-40]. In particular, it is typical for people suffering from depression to talk about their illness and the drugs they are taking [41-43].

Twitter is a very popular microblogging platform with more than 330 million active users worldwide [44]. Tweets, freely available in almost 90% of users' accounts, provide a huge amount of data that can collected in real time [28,30,33,45-48]. Twitter users post short messages about facts, feelings, and opinions, including about health conditions [49].

Mining of drug-related information from Twitter has been applied in the pharmacovigilance field [27,50]. Some pharmacovigilance studies carried out on Twitter studied specific cohorts by identifying users' mentions of drug intake [37,51-53]. Other studies focused on adverse drug reactions, analyzing users' tweets regarding adverse events and side effects associated with drug use, which were identified by means of generic or brand names [29,47,54,55]. In our previous study [49], we observed that Twitter users who are potentially suffering from depression show particular behavioral and linguistic features in their tweets. These features were related to an increase in their activity during the night, a different style of writing with increased use of the first-person singular pronoun, fewer characters in their tweets, an increase in the frequency of words related to sadness and disgust emotions, and more frequent presence of negation words and negative polarity. This information can be used as a complementary tool to detect signals of depression and for monitoring and supporting patients using Twitter.

Objectives

In this paper, we aim to enrich our previous study [49] by focusing on analysis of the changes in behavioral and linguistic features of Twitter users in Spanish language, which may be associated with the antidepressant medication these users are taking. It is worth mentioning that users from Spanish-speaking countries are among the most active on Twitter in the world [56]. The study is focused on Twitter users who mention treatment with SSRIs, which are the most frequently prescribed antidepressants [15]. In particular, this study compares the characteristics of the tweets posted while users were probably taking SSRIs versus the tweets posted by the same users when they have a lower probability of taking this antidepressant medication. This analysis can contribute to better understanding how these drugs affect users' mood. Although we found two additional studies describing changes in Twitter users' language in some mental disorders [57,58], to the best of our knowledge, there are no other studies that analyze Twitter posts in Spanish language to detect behavioral and linguistic changes when the users are taking antidepressant medication.

Methods

Study Design

This study was designed with the aim of analyzing the behavioral patterns and linguistic features of users who mention SSRIs in their Twitter timeline. The study was developed in several steps and focused on tweets written in Spanish. The flow diagram of the study is depicted in Figure 1.

As shown in Figure 1, two nonoverlapping datasets of tweets from users mentioning treatment with SSRIs were obtained: (1) The *in-treatment tweets dataset* was made up of the tweets posted throughout the 30 days after the publication date of any tweet mentioning SSRI intake. We assumed that these tweets were posted while the users had a high probability of being in treatment with an SSRI. (2) The *unknown-treatment tweets dataset* was made up of the tweets that were posted more than 90 days before or more than 90 days after the publication date of any tweet mentioning SSRI intake. We assumed that these tweets were posted while users had a lower probability of being in treatment with an SSRI than in the previous dataset.

These datasets were designed in a way that made it possible to carry out intrasubject comparisons, since the *in-treatment* tweets and *unknown-treatment* tweets datasets were obtained from the same Twitter users.

The strategy for the selection of the tweets included in the two datasets is depicted in Figure 2.

Figure 1. Flow diagram of the study process. SSRI: selective serotonin reuptake inhibitor.

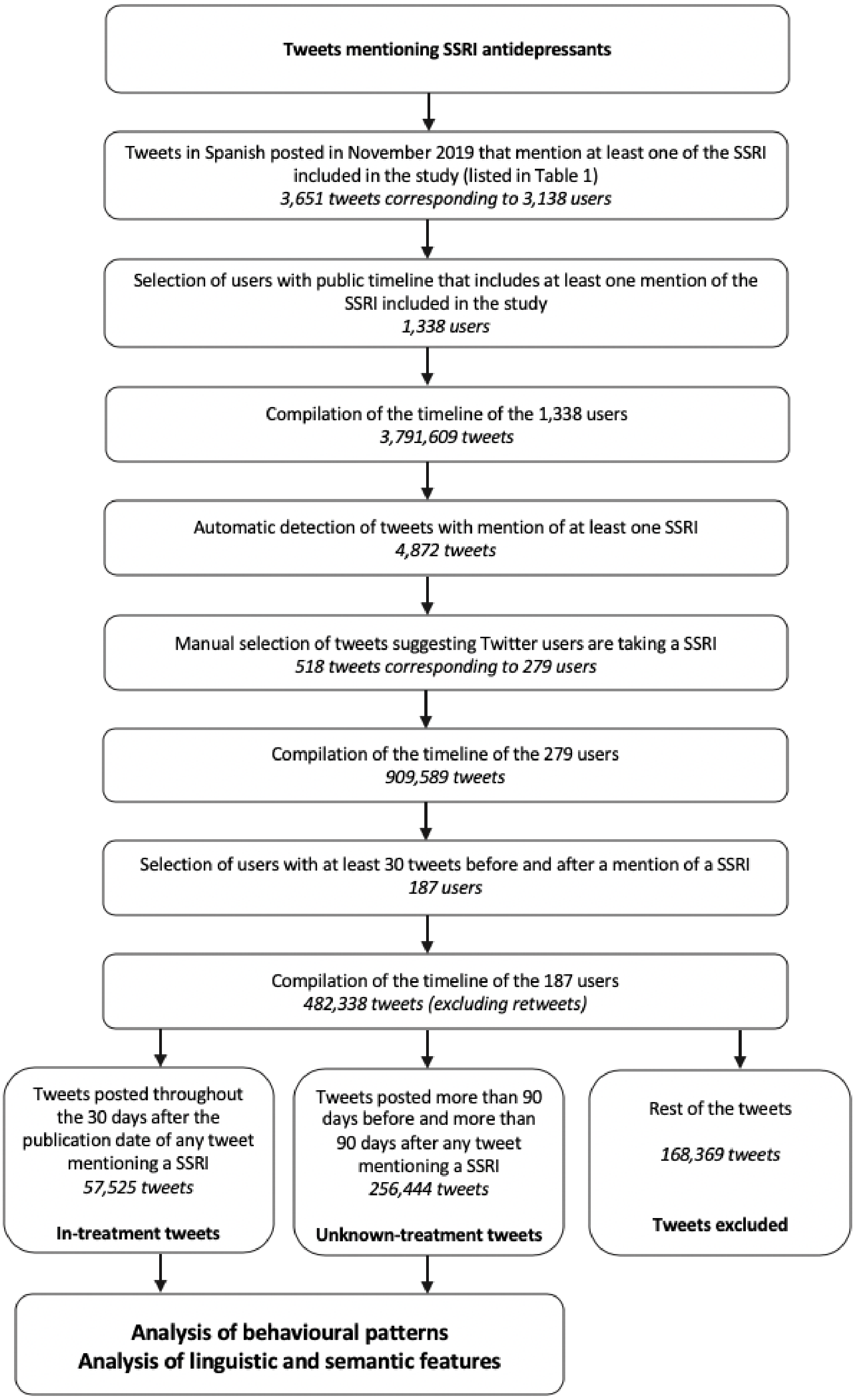
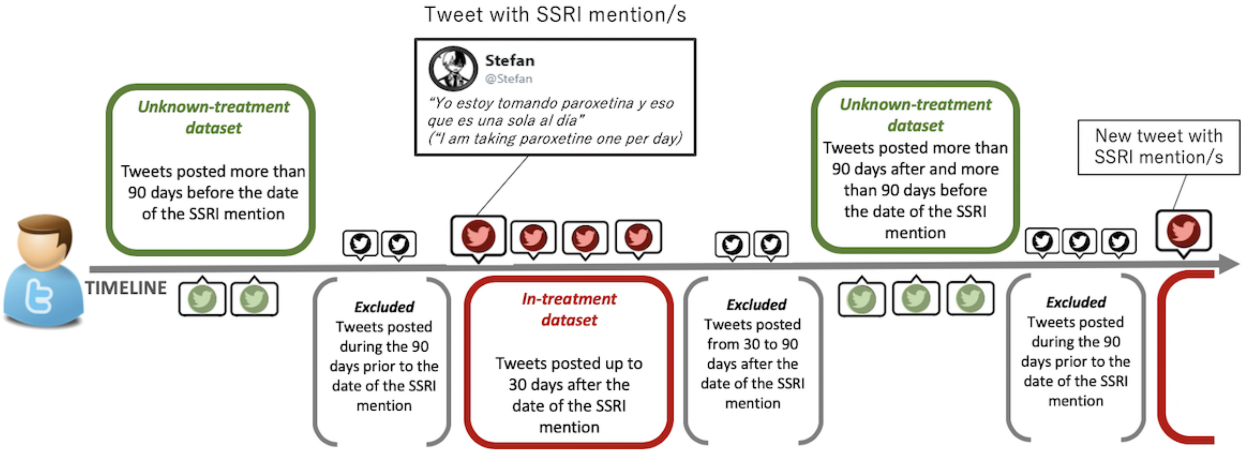


Figure 2. The in-treatment and unknown-treatment dataset selection strategy. SSRI: selective serotonin reuptake inhibitor.



Data Collection and User Selection

The selection of the tweets and their users was based on the filtered real-time streaming support provided by the Twitter application programming interface [59]. In the first step, we selected tweets in Spanish that mention any of the SSRI generic and brand names used around the world. To obtain the generic and brand names, we performed searches on the following databases and resources: DrugBank [60], the Anatomical Therapeutic Chemical Classification System and the Defined Daily Dose of the World Health Organization [61], Wikipedia [62], and the Database for Pharmacoepidemiological Research in Primary Care [63]. The list of 135 generic and brand names obtained is shown in Table 1.

Table 1. Selective serotonin reuptake inhibitors (SSRIs) used in the study.

Generic name	Brand names
Fluvoxamina (fluvoxamine)	Dumirox, Faverin, Floxyfral, Fluvoxin, Luvox, Uvox
Fluoxetina (fluoxetine)	Prozac, Reneuron, Adofen, Luramon, Sarafem
Paroxetina (paroxetine)	Seroxat, Motivan, Frosinor, Praxil, Daparox, Xetin, Apo-oxpar, Appoxar, Aropax, Aroxat, Aroxat CR, Bectam, Benepax, Casbol, Cebrilin, Deroxat, Hemtrixil, Ixicrol, Loxamine, Meplar, Olane, Optipar, Oxetine, Pamax, ParadiseCR, Paradox, Paraxyle, Parexis, Paroxat, Paroxet, Paxan, Paxera, Paxil, Paxil CR, Pexot, Plasare, Pondera, Posivyl, Psicoasten, Rexetin, Seretran, Sereupin, Tiarix, Tamcere, Traviata, Xerenex, Xetroran
Sertralina (sertraline)	Aremis, Besitran, Zoloft, Altisben, Aserin, Altruline, Ariale, Asertral, Atenix, Eleval, Emergen, Dominium, Inosert, Irradial, Sedora, Serolux, Sertex
Citalopram (citalopram)	Seropram, Celexa, Akarin, C Pram S, Celapram, Celica, Ciazil, Cilate, Cilift, Cimal, Cipralex, Cipram, Cipramil, Cipraned, Cinapen, Ciprapine, Ciprotan, Citabax, Citaxin, Citalec, Citalex, Citalo, Citalopram, Citol, Citox, Citrol, Citta, Dalsan, Denyl, Elopram, Estar, Humorup, Humorap, Oropram, Opra, Pram, Pramcit, Procimax, Recital, Sepram, Szetalo, Talam, Temperax, Vodelax, Zentius, Zetalo, Cipratal, Zylotex
Escitalopram (escitalopram)	Cipralex, Diprex, Esertia, Essential, Heipram, Lexapro

The following 7 brand names of medicines have been excluded due to their semantic ambiguity: Essential, Motivan, Estar, Traviata, Pondera, Recital, and Emergen. These commercial names are, at the same time, very common words used with different meanings in Spanish, as we verified after reviewing a random sample of 200 tweets with mentions of these words. The number of tweets excluded because of their semantic ambiguity was 21,104. In the manual check of a random sample of 200 tweets, the mentions of SSRIs when using these words were 0% (0/200) in some cases, such as for Motivan and Estar, and 0.5% (1/200) for Recital. The final list of words included 128 generic and brand names of SSRIs.

Using the aforementioned 128 SSRI names, we collected 3651 tweets in Spanish posted during November 2019 with occurrences of the words listed in Table 1. These tweets were posted by 3138 different Twitter users and mentioned 33 different words from the list. The frequencies of these 32 words are shown in Table 2.

Table 2. Frequencies of SSRI names mentioned in Spanish tweets during November 2019.

SSRI mentions	Frequency
Prozac	998
Fluoxetina	756
Sertralina	542
Escitalopram	248
Citta	210
Citalo	109
Paroxetina	69
Pram	49
Fluvoxamina	40
Citalopram	33
Seroxat	22
Eleval	21
Lexapro	20
Opra	18
Casbol	14
Ariale	11
Zoloft	9
Altruline	9
Paxil	7
Akarin	7
Heipram	4
Aremis	4
Cimal	3
Tiarix	2
Seretran	2
Dominium	2
Citox	2
Atenix	2
Aserin	2
Talam	1
Dalsan	1
Celexa	1

In a second step, we crawled the public Twitter timelines of the 3138 users (until the 3200 most recent tweets for each user were retrieved). Given that retweets are not useful for analyzing the linguistic behavior of a particular user, the third step consisted of excluding the retweets and checking if the remaining tweets from each timeline included the mention of at least one SSRI. 1800 users were excluded by this filter, leaving a total of 1338 Twitter users. We obtained 3,791,609 tweets after compiling the timelines from these 1338 users. From these timelines, 4872 tweets mentioning at least one of the SSRIs from the list were automatically detected. These 4872 tweets were independently reviewed by two experts, a psychologist and a family physician, both with clinical experience. These experts manually selected the tweets that suggested that the user who posted the tweet was taking an SSRI on the date of posting. Examples of these tweets are shown in Textbox 1.

Textbox 1. Examples of tweets that positively or negatively suggest whether the user is taking an SSRI.

Positive examples:

- "Eso de tener sueños raros debido a la fluoxetina se está saliendo de control." ("Having odd dreams due to fluoxetine is getting out of control.")
- "Yo tomo sertralina, como me lo receta el doctor y aún así a veces siento que el mundo donde estoy no es para mi. Ese susto esa angustia esas ganas de correr es algo que sólo el que lo padece lo entiende" ("I take sertraline as my doctor prescribes it to me and, even so, sometimes I feel that the world I'm living in is not for me. This fear this anxiety this desire to run out is something that only one who suffers from it can understand")

Negative examples:

- "Ella debería tomar prozac, como Tic Tac" ("She should take prozac, like Tic Tac" [a candy brand])
- "La Paroxetina es un medicamento que pertenece a la familia de los antidepresivos inhibidores de la recaptación de la serotonina ¡Conoce más sobre él!" ("Paroxetine is a drug that belongs to the antidepressant family of serotonin reuptake inhibitors. Find out more about it!")

The agreement between reviewers was 93.1% (4537/4872) with a Cohen kappa score of 0.68, indicating that there was substantial agreement between raters. The reviewers discussed and reached a consensus on the classification of the 335 tweets they classified differently. Finally, we obtained a total number of 518 tweets with one or more SSRI mentions, suggesting that the users who posted these tweets were taking an SSRI at the moment of posting. These tweets corresponded to 279 different users. Therefore, these users had two characteristics: first, the tweets on their timeline included at least one mention of SSRIs, and second, the text of tweets mentioning SSRIs suggested that the user was taking the antidepressant. In addition, we analyzed the tweets posted by each user that belonged to the two datasets (*in-treatment* and *unknown-treatment*; see Figure 1) by trying different minimum numbers of tweets per dataset (10, 30, 60, and 100 tweets) in order to include a user in the study. 10 tweets contained little information in terms of number of words or posting characteristics. In the cases of 60 and 100 tweets, the number of users included dropped dramatically. For this reason, we applied a requirement of a minimum of 30 tweets in both *in-treatment* and *unknown-treatment* datasets to keep the balance between the number of tweets and the number of users to be included in the study. After applying this requirement, 187 users were finally included in the study. The complete timelines of these users were compiled, totaling 668,842 tweets, which were reduced to 482,338 once retweets were removed. Out of these, 168,369 more tweets were excluded because they were posted on dates located outside the periods that qualified a tweet for being included in the *in-treatment* or the *unknown-treatment* datasets. Finally, 57,525 tweets were included in the *in-treatment* dataset and 256,444 in the *unknown-treatment* dataset.

Data Analysis

The two datasets of tweets, *in-treatment* and *unknown-treatment*, were compared in order to determine the existence of behavioral and linguistic differences between the tweets generated by the users in each period. The features that were analyzed are listed in Table 3.

Table 3. Features of the tweets analyzed.

Features	Analyses performed
Distribution over time	Tweets per hour, tweets during daytime vs night, tweets per day, tweets during weekdays vs weekend
Length	Number of characters, number of words
Part-of-speech (POS)	Number of words by grammatical categories (part-of-speech tags)
Emotion analysis	Frequencies of emotion types
Negations	Frequencies of negation words
Polarity	Polarity of tweets on the basis of Spanish Sentiment Lexicon

Paired data statistical significance tests (paired *t* tests) were carried out whenever possible. The Benjamini-Hochberg false discovery rate was applied for multiple testing correction analysis [64]. The *P* values provided incorporate it.

The textual content of each tweet was analyzed using the same methodology and tools used in our previous study [49]. The textual content of each tweet was analyzed by means of the following steps: tokenization performed based on a customized Twitter tokenizer included in the Natural Language Toolkit [65]; part-of-speech (POS) tagging performed by means of the FreeLing Natural Language Processing tool in order to analyze the usage patterns of grammatical categories, such as verbs, nouns, pronouns, adverbs, and adjectives, in the text of tweets [66]; identification of negations performed by building upon a customized list of Spanish negation expressions, such as nada (nothing), nadie (nobody), no (no), nunca (never), and similar; identification of positive and negative words inside the text of each tweet using the Spanish Sentiment Lexicon [67]; and identification of words and expressions associated with emotions such as happiness, anger, fear, disgust, surprise, and sadness [68] by using the Spanish Emotion Lexicon [69].

The statistical analyses were carried out using Python 3.7, the Tweepy, SciPy, and Natural Language Toolkit libraries, and R version 3.6.2 (R Development Core Team), including the R "psych" package 1.9.12.31. All the aforementioned software tools are publicly available.

Ethical Approval

The protocol used in this study was reviewed and approved by the Ethics Committee of Parc Salut Mar (approval number 2017/7234/1).

Results

Distribution Over Time

Several types of distribution-over-time analysis were performed in order to study the potential influence of being in *in-treatment* periods in comparison to *unknown-treatment* ones. The tweet hours were adjusted by the users' time zone.

The mean duration of the time period analyzed of all the users was 28.2 months (SD 24.7); the mean of the total number of tweets analyzed was 307.6 (SD 336.0) for *in-treatment* periods and 1371.4 (SD 748.2) in the case of *unknown-treatment* periods. The mean number of tweets per day generated by users during *in-treatment* periods was 11.44 (SD 10.05); this number dropped to 9.07 (SD 7.21) in the *unknown-treatment* dataset with a mean difference of 2.37 (SD 9.72) between periods, which shows statistically significant differences between the two datasets (t_{186}=3.33; P<.001).

The mean percentage of tweets posted during daytime (between 8 AM and midnight) was 64.30% (SD 14.83) when the users were *in-treatment* periods; this percentage fell to 61.78% (SD 13.69) during the *unknown-treatment* periods, with a mean percentage difference of 2.52% (SD 11.81), which implies statistically significant differences (t_{186}=3.07; P=.004).

The mean number of tweets generated during the weekdays (from Monday to Friday) was 12.28 (SD 11.05) during *in-treatment* periods and 9.33 (SD 6.70) in the *unknown-treatment* periods, with a mean difference of 2.95 (SD 10.23) and statistically significant differences between the datasets (t_{186}=3.93; P<.001). For the mean number of tweets generated during the weekends (Saturday and Sunday), it was 9.35 (SD 9.31) in the *in-treatment* period and 8.41 (SD 9.82) in the *unknown-treatment* period, with a mean difference of 0.94 (SD 10.92) that implies statistically significant differences between the datasets (t_{186}=1.18; P=.23). The mean percentage of tweets posted on weekdays was 75.95% (SD 9.17) during *in-treatment* periods; the percentage went down to 74.40% (SD 5.31) in *unknown-treatment* periods, with a mean percentage difference of 1.56% (SD 8.9) that implies statistically significant differences between the two periods (t_{186}=2.39; P=.02).

Length

The average number of characters per tweet was 88.03 (SD 30.74) and 85.19 (SD 28.82) in the *in-treatment* and *unknown-treatment* datasets, respectively, with a mean difference of 2.84 (SD 17.70) and statistically significant differences between the periods (t_{186}=2.19; P=.03). As for the number of words per tweet, the mean was 15.68 (SD 5.75) in the *in-treatment* dataset and 15.09 (SD 5.20) in the *unknown-treatment* dataset, with a mean difference of 0.59 (SD 3.54) and statistically significant differences (t_{186}=2.28; P=.02).

Links and Mentions to Other Users

The mean percentages of tweets that include at least one link were 23.10% (SD 16.16) and 23.27% (SD 15.29) in the *in-treatment* and *unknown-treatment* datasets, respectively, with a mean difference of −0.17 (SD 10.94), which is not statistically significant (t_{186}=−0.23; P=.82).The mean percentages of tweets that include at least one mention of another Twitter user were 45.79% (SD 24.77) and 43.52% (SD 24.71) in the *in-treatment* and *unknown-treatment* datasets, respectively, with a mean difference of 2.27% (SD 12.13), which is statistically significant (t_{186}=2.56; P=.01).

Part-of-Speech

As for the analysis of the number of words by grammatical category (ie, part-of-speech) in each tweet, we also compared the *in-treatment* and *unknown-treatment* datasets. The mean percentage of words per grammatical category over the total number of words in each dataset is shown in Table 4. We considered the most relevant lexical POS such as verbs, nouns, pronouns, adverbs, and adjectives, excluding conjunctions, interjections, punctuations, determiners, adpositions, numbers, and dates.

Regarding the different types of pronouns, the mean percentages of personal pronouns in each dataset are shown and compared in Table 5.

Table 4. Percentages of part-of-speech words compared between in-treatment and unknown-treatment datasets.

POS[a]	*in-treatment* (%), mean	*unknown-treatment* (%), mean	Difference (%), mean (SD)	Paired *t* test	*P* value
Verbs	18.50	18.20	0.3 (1.28)	3.15	.002
Nouns	19.50	19.94	−0.44 (2.57)	−2.35	.02
Pronouns	9.19	8.93	0.26 (1.33)	2.61	.01
Adverbs	6.42	6.36	0.06 (0.84)	0.97	.34
Adjectives	6.05	6.21	−0.16 (0.95)	−2.34	.02

[a] POS: part-of-speech.

Table 5. Mean percentages of personal pronouns compared between in-treatment and unknown-treatment datasets.

Personal pronouns	*in-treatment* (%), mean	*unknown-treatment* (%), mean	Difference (%), mean (SD)	Paired *t* test	*P* value
1st person singular	49.50	47.80	1.7 (8.68)	2.67	.008
2nd person singular	14.77	16.07	−1.3 (6.17)	−2.88	.004
3rd person singular	22.13	22.86	−0.73 (5.79)	−1.72	.08
1st person plural	3.44	3.43	0.01 (3.43)	0.04	.96
2nd person plural	1.00	1.00	0 (1.22)	−0.01	.98
3rd person plural	5.60	5.39	0.21 (3.68)	0.77	.44

Emotion Analysis

The mean percentages of the different emotions, obtained using the Spanish Sentiment Lexicon on the tweets posted in the two periods, are shown in Table 6.

Table 6. Mean percentages of different emotions compared between in-treatment and unknown-treatment datasets.

Emotion	*in-treatment* (%), mean	*unknown-treatment* (%), mean	Difference (%), mean (SD)	Paired *t* test	*P* value
Happiness	26.93	25.94	0.99 (5.82)	2.32	.02
Sadness	10.01	9.76	0.25 (4.20)	0.81	.41
Fear	3.20	3.02	0.18 (1.94)	1.23	.21
Anger	5.52	5.20	0.32 (2.71)	1.62	.11
Disgust	3.11	3.06	0.05 (1.97)	0.38	.69
Surprise	5.59	5.06	0.53 (2.42)	2.98	.003

Negation Analysis

The mean percentages of tweets, among all users, that included one or more negation words were 27.66% (SD 10.54) and 26.59% (SD 9.87) for the *in-treatment* and *unknown-treatment* datasets, respectively, with a mean difference of 1.07% (SD 6.99), which is statistically significant (t_{186}=2.10; *P*=.04).

Polarity Analysis

As for the polarity of tweets, the percentage of tweets, among all users, with one or more positive words inside the text was 15.13% (SD 6.56) in the *in-treatment* dataset and 14.50% (SD 5.43) in the *unknown-treatment* dataset, with a mean percentage difference of 0.63% (SD 5.22; t_{186}=1.66; *P*=.09). The percentage of tweets with one or more negative words was 7.97% (SD 4.40) in the *in-treatment* dataset and 7.54% (SD 3.52) in the *unknown-treatment* dataset, with a mean percentage difference of 0.43% (SD 3.58) (t_{186}=1.64; *P*=.10). No statistically significant differences were detected in this analysis.

Discussion

Principal Findings

Social media platforms in general, and Twitter in particular, may provide useful information on how patients respond when they receive a pharmacological treatment, as has been shown in several studies in which social media has been used as a complementary source of pharmacovigilance and monitoring [34,70]. In this study, we analyzed the tweets of users who mentioned they were taking antidepressant drugs, in particular SSRIs, with the aim of detecting behavioral changes when they are more likely to be in treatment in comparison to periods in which they are less likely to be in treatment (*"in-treatment"* vs *"unknown-treatment"* periods).

The results of this study show that, in general, Twitter users significantly increased their activity of posting tweets during the *in-treatment* periods. This increase was more pronounced during weekdays than during weekends. We also observed a significantly greater proportion of tweets posted during the daytime during the *in-treatment* periods. These results are consistent with the results of our previous paper [49], in which we observed that the control group without signs of depression showed more tweet posting activity than the group of users with signs of depression, especially during the daytime and the weekdays. These results are also consistent with another paper that described the behavior in social media of people with self-reported depression [41], as well as with a study on the diurnal mood variation of patients suffering from major depressive disorder [71]. In summary, we can state that when considering tweet posting activity, the behavior of individuals suffering from depression becomes more similar to that of the general population when they are in treatment with SSRIs.

Likewise, the average number of characters and words per tweet were significantly higher when the Twitter users were in treatment with SSRIs, a finding that again points toward an increase in the activity of these treated users. In addition, the increase in the number of mentions per tweet can reflect a greater interest in interacting with other people. All these changes may be due to some improvement in their anhedonic symptoms because of the medication.

Regarding the linguistic analysis, we observed quantitatively slight changes between the *in-treatment* and the *unknown-treatment* periods, although in some cases they are statistically significant. These slight findings are not easily interpretable. In general, given that the style of writing of people suffering from depression is characterized by self-focus attention, which is associated with negative emotional states and psychological distancing in order to connect with others [72], we can conclude that when the studied subjects were in treatment, they improved some traits related to their posting activity as previously mentioned, but at the same time, their language maintained the features of people suffering from depression without a clear influence of the medication.

Emotion is another important aspect that characterizes people suffering from depression, and it was consequently analyzed. When the users were in treatment, they showed small but statistically significant increases in the happiness and surprise emotions, but not in sadness or other emotions (ie, anger, fear and disgust). As for the number of negations, the users slightly increased their use of these types of words during the *in-treatment* period. However, the polarity analysis did not show differences between the periods.

The increased activity observed on Twitter when the users were likely to be in treatment with SSRIs can be linked to improved emotional status in their happiness and surprise emotions. These changes are consistent with our previous observations on mood states of Twitter users without depression compared to those with depression [49]. However, the traits that are related to language, as indicated by the POS analysis and the use of negations, maintained a similar profile to that of subjects with depression, independently of the pharmacological treatment detected. These results denote that users with depression who are taking SSRIs show some mood improvements while receiving antidepressant treatment, but at the same time maintain an altered language pattern, which may be indicative of incomplete recovery.

On the basis of our statistically significant results, we may state that Twitter timelines can be used as a complementary tool to monitor subjects in order to detect adherence to treatment, which is an important problem in this kind of patient. Adherence to treatment is essential for disease remission [73-76]. According to some studies, it is common for patients suffering from depression to not maintain the duration of antidepressant treatment that is clinically recommended [4,18,77]. In summary, the follow-up of behavioral and language changes in users' Twitter timelines can be useful for monitoring the evolution of depressive symptoms and the effect of treatments.

Limitations and Future Directions

This type of study in general, and this one in particular, presents some limitations. For instance, we considered tweets written in Spanish and from public Twitter users' timelines, and these users may be not representative of the general population or people suffering from depression [33,49,78,79]. Some studies have shown that Twitter users are often urban people with high levels of education, and they are generally younger than the general population [33,49,78,80,81]. We should also take into account that SSRIs are used in different types of depressive disorders and in other mental conditions. Moreover, we have no information about whether these drugs were taken in the context of a prescribed medical treatment or as a result of an inappropriate self-medication decision.

Another limitation may be the fact that Twitter users who share their personal drug intake may use words or expressions not included in the list of drug names employed in this study for streaming tweets, even though we tried to be exhaustive in the list of names used. Twitter texts are informal and limited by the number of characters, and they commonly include abbreviations, errors, or slang language [33,45]. All these issues can make it difficult to automatically extract drug mentions and link them to a formal lexicon [28,30,50,53,55]. Unlike clinical records that could be linked to domain resources, the lack of lay vocabularies related to health concepts and terminologies hinders the processing of social media texts [55]. In addition, the results obtained may depend on the particular drugs selected for the study [33], as well as on the periods of time set up for classifying the tweets into the *in-treatment* and *unknown-treatment* datasets. On the basis of the strategy applied for defining the groups of tweets to be compared (tweets generated just after mentions to SSRI intake vs tweets generated in periods far from any mention to the SSRI intake), there is some chance of misclassification; it is likely that not all the tweets in the first group were generated by users under actual SSRI treatment, and it is probable that some tweets of the second group have been generated by users under SSRI treatment.

Furthermore, we must take into account that data from social media posts contain irrelevant information. Although the proportion of useful information for the specific research purpose can be quite limited, it constitutes a useful starting point [28,30,51,53]. In this scenario, the human curation of tweets is a necessary step in this kind of analysis [34]. Even so, due to the different nuances that a tweet can involve, it is not easy to detect real drug intakes or firsthand experiences [24,46,52].

Conclusions

Social media can be used to monitor the health status of people and, in particular, to detect symptoms or features related to diseases or health conditions by means of analysis of the users' behavior and language on social media platforms. Moreover, the detection of changes in symptoms or other features when patients are taking medications can provide interesting insights for monitoring pharmacological treatments, as well as for following up on the evolution of the disease, detecting side effects, or providing information related to treatment adherence. Changes in some features, such as a decrease in activity on Twitter or of the length of tweets, an increase of self-focus through the use of the first-person singular pronoun, and changes in the happiness and surprise emotions could be used as complementary tools to detect the worsening of the psychological status of users suffering from depression, as well as to perceive lack of adherence to treatment. This information may be especially useful in patients suffering from chronic diseases who are receiving long-term treatments, as is the case for mental disorders in general and depression in particular. However, it is not possible to determine the specific reasons why individuals change their behavior and language on social media platforms in the framework of a disease and its treatment without performing a clinical assessment. Overall, this study shows the relevance of monitoring behavioral and linguistic changes in the tweets of persons taking antidepressants. These changes are likely to be influenced by the diverse stages of the disease and the therapeutic effects of the treatment that these Twitter users are receiving, opening a new line of research to better understand and follow up on depression through social media.

Acknowledgments

We received support from the Agency for Management of University and Research Grants in Catalonia (Spain) for the incorporation of new research personnel (FI2016) and from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement number 802750 (FAIRplus) with the support of the European Union’s Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations Companies.

The Research Programme on Biomedical Informatics is a member of the Spanish National Bioinformatics Institute, funded by Instituto de Salud Carlos III and the European Regional Development Fund (PRB2-ISCIII), and it is supported by grant PT17/0009/0014. The Department of Experimental and Health Sciences, Universitat Pompeu Fabra, is a “Unidad de Excelencia María de Maeztu”, funded by the Ministry of Economy, Spain [MDM-2014-0370]. Funding for the open access charge is from the Agència de Gestió d’Ajuts Universitaris i de Recerca Generalitat de Catalunya (2017 SGR 00519).

The Database for Pharmacoepidemiological Research in Primary Care, from the Spanish Agency for Medicines and Health Products of the Ministry of Health and Consumer Affairs and Social Welfare of the Government of Spain, was used to obtain useful information about the prescription frequency of antidepressants.

Conflicts of Interest

None declared.

References

1. World Health Organization. Depression: Key Facts. 2019. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2020-01-09]
2. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. 2017. URL: http://apps.who.int/iris/bitstream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf?ua=1 [accessed 2020-01-09]
3. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013 Nov;10(11):e1001547 [FREE Full text] [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]
4. Royal College of Psychiatrists. Position statement on antidepressants and depression. PS04/19. London, United Kingdom: Royal College of Psychiatrists; May 2019.
5. Marcus M, Yasamy MT, van Ommeren M, Chisholm D, Saxena S, WHO Department of Mental Health and Substance Abuse. Depression: a global public health concern.: World Health Organization; 2012. URL: http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf [accessed 2020-11-12]
6. Global BODS2C. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the global burden of disease study 2013. Lancet Aug 22 2015;386(9995):743-800. [Medline: 26063472]
7. Trautmann S, Rehm J, Wittchen H. The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? EMBO Rep 2016 Sep;17(9):1245-1249 [FREE Full text] [doi: 10.15252/embr.201642951] [Medline: 27491723]
8. Patel V, Chisholm D, Parikh R, Charlson F, Degenhardt L, Dua T. Global priorities for addressing the burden of mental, neurological, substance use disorders. In: Patel V, Chisholm D, Dua T, Laxminarayan R, Medina-Mora ME, Vos T. eds. Disease Control Priorities: Mental, Neurological, and Substance Use Disorders, 3rd ed. (Vol 4). Washington, DC: The World Bank; 2016:1-27.
9. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov 9;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
10. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: Systematic review. J Med Internet Res 2017 Dec 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
11. Working Group of the Clinical Practice Guideline on the Management of Depression in adults. Ministry of Health, Social Services and Equality. Galician Agency for Health Technology Assessment (avalia-t). Clinical practice guideline on the management of depression in adults. 2014. URL: https://portal.guiasalud.es/wp-content/uploads/2018/12/GPC_534_Depresion_Adulto_Avaliat_compl_en.pdf [accessed 2020-11-12]

12. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
13. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
14. Kraus C, Kadriu B, Lanzenberger R, Zarate CA, Kasper S. Prognosis and improved outcomes in major depression: a review. Transl Psychiatry 2019 Apr 03;9(1):127 [FREE Full text] [doi: 10.1038/s41398-019-0460-3] [Medline: 30944309]
15. Ministerio de Sanidad, Servicios Sociales e Igualdad. Agencia Española de Medicamentos y Productos Sanitarios (AEMPS). Utilización de medicamentos antidepresivos en España durante el periodo 2000-2013. Informe de utilización de medicamentos U/AD/V1/14012015. Madrid: Ministerio de Sanidad, Servicios Sociales e Igualdad. Agencia Española de Medicamentos y Productos Sanitarios (AEMPS); 2015.
16. Luo Y, Kataoka Y, Ostinelli EG, Cipriani A, Furukawa TA. National prescription patterns of antidepressants in the treatment of adults with major depression in the US between 1996 and 2015: A population representative survey based analysis. Front Psychiatry 2020;11:35 [FREE Full text] [doi: 10.3389/fpsyt.2020.00035] [Medline: 32116850]
17. Fasipe OJ. The emergence of new antidepressants for clinical use: Agomelatine paradox versus other novel agents. IBRO Rep 2019 Jun;9(6):95-110 [FREE Full text] [doi: 10.1016/j.ibror.2019.01.001] [Medline: 31211282]
18. Depression in adults: recognition and management. Clinical guideline CG90. National Institute for Health and Care Excellence (NICE); 2009 Oct 28. URL: https://www.nice.org.uk/guidance/cg90 [accessed 2020-11-12]
19. Gelenberg A, Freeman M, Markowitz J, Rosenbaum J, Thase M, Trivedi M, et al. Practice guideline for the treatment of patients with major depressive disorder. 3rd ed. Washington, DC: American Psychiatric Association; 2010.
20. Lane R, Baldwin D, Preskorn S. The SSRIs: advantages, disadvantages and differences. J Psychopharmacol 1995 Jan;9(2 Suppl):163-178. [doi: 10.1177/0269881195009002011] [Medline: 22297235]
21. Pundiak TM, Case BG, Peselow ED, Mulcare L. Discontinuation of maintenance selective serotonin reuptake inhibitor monotherapy after 5 years of stable response: a naturalistic study. J Clin Psychiatry 2008 Nov;69(11):1811-1817. [doi: 10.4088/jcp.v69n1117] [Medline: 19026252]
22. Emslie GJ, Mayes TL, Ruberu M. Continuation and maintenance therapy of early-onset major depressive disorder. Paediatr Drugs 2005;7(4):203-217. [doi: 10.2165/00148581-200507040-00001] [Medline: 16117558]
23. Garnock-Jones KP, McCormack PL. Escitalopram: a review of its use in the management of major depressive disorder in adults. CNS Drugs 2010 Sep;24(9):769-796. [doi: 10.2165/11204760-000000000-00000] [Medline: 20806989]
24. Clevenger SS, Malhotra D, Dang J, Vanle B, IsHak WW. The role of selective serotonin reuptake inhibitors in preventing relapse of major depressive disorder. Ther Adv Psychopharmacol 2018 Jan;8(1):49-58 [FREE Full text] [doi: 10.1177/2045125317737264] [Medline: 29344343]
25. Sim K, Lau WK, Sim J, Sum MY, Baldessarini RJ. Prevention of relapse and recurrence in adults with major depressive disorder: Systematic review and meta-analyses of controlled trials. Int J Neuropsychopharmacol 2015 Jul 07;19(2):pyv076 [FREE Full text] [doi: 10.1093/ijnp/pyv076] [Medline: 26152228]
26. Peselow ED, Tobia G, Karamians R, Pizano D, IsHak WW. Prophylactic efficacy of fluoxetine, escitalopram, sertraline, paroxetine, and concomitant psychotherapy in major depressive disorder: outcome after long-term follow-up. Psychiatry Res 2015 Feb 28;225(4):680-686. [doi: 10.1016/j.psychres.2014.11.022] [Medline: 25496869]
27. Salathé M. Digital pharmacovigilance and disease surveillance: Combining traditional and Big-Data systems for better public health. J Infect Dis 2016 Dec 01;214(suppl_4):S399-S403 [FREE Full text] [doi: 10.1093/infdis/jiw281] [Medline: 28830106]
28. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
29. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying adverse effects of HIV drug treatment and associated sentiments using Twitter. JMIR Public Health Surveill 2015;1(2):e7 [FREE Full text] [doi: 10.2196/publichealth.4488] [Medline: 27227141]
30. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]
31. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018 Jan 04;14(1):1 [FREE Full text] [doi: 10.1186/s40504-017-0065-7] [Medline: 29302758]
32. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
33. Lardon J, Bellet F, Aboukhamis R, Asfari H, Souvignet J, Jaulent M, et al. Evaluating Twitter as a complementary data source for pharmacovigilance. Expert Opin Drug Saf 2018 Aug;17(8):763-774. [doi: 10.1080/14740338.2018.1499724] [Medline: 29991282]
34. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. Drug Saf 2014 May;37(5):343-350 [FREE Full text] [doi: 10.1007/s40264-014-0155-x] [Medline: 24777653]

35. Crannell WC, Clark E, Jones C, James TA, Moore J. A pattern-matched Twitter analysis of US cancer-patient sentiments. J Surg Res 2016 Dec;206(2):536-542. [doi: 10.1016/j.jss.2016.06.050] [Medline: 27523257]
36. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. J Med Internet Res 2016;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
37. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
38. Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Oct 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
39. Park M, Cha C, Cha M. Depressive moods of users portrayed in Twitter. In: Proceedings of the ACM SIGKDD Workshop on Health Informatics. 2012 Presented at: HI-KDD'12; August 12-16; Beijing, China p. 1-8.
40. Conway M, O'Connor D. Social Media, Big Data, and Mental Health: Current Advances and Ethical Implications. Curr Opin Psychol 2016 Jun;9:77-82 [FREE Full text] [doi: 10.1016/j.copsyc.2016.01.004] [Medline: 27042689]
41. De Choudhury C, Gamon M, Counts S, Horvitz E. Predicting depression via social media. In: Proceedings of the Seventh International Conference on Weblogs and Social Media. 2013 Presented at: AAA'13; July 8-11; Cambridge, MA p. 128-138.
42. Cavazos-Rehg P, Krauss M, Sowles S, Connolly S, Rosas C, Bharadwaj M, et al. A content analysis of depression-related tweets. Comput Human Behav 2016 Jan 01;54:351-357 [FREE Full text] [doi: 10.1016/j.chb.2015.08.023] [Medline: 26392678]
43. Nguyen T, O'Dea B, Larsen M, Phung D, Venkatesh S, Christensen H. Using linguistic and topic analysis to classify sub-groups of online depression communities. Multimed Tools Appl 2015 Dec 21;76(8):10653-10676. [doi: 10.1007/s11042-015-3128-x]
44. Statista. Number of Monthly Active Twitter Users Worldwide From 1st Quarter 2010 to 2nd Quarter 2019 (in Millions). 2019. URL: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ [accessed 2020-02-03]
45. Audeh B, Calvier F, Bellet F, Beyens M, Pariente A, Lillo-Le Louet A, et al. Pharmacology and social media: Potentials and biases of web forums for drug mention analysis-case study of France. Health Informatics J 2019 Sep 30;26(2):1253-1272. [doi: 10.1177/1460458219865128] [Medline: 31566468]
46. Alvaro N, Conway M, Doan S, Lofi C, Overington J, Collier N. Crowdsourcing Twitter annotations to identify first-hand experiences of prescription drug use. J Biomed Inform 2015 Dec;58:280-287 [FREE Full text] [doi: 10.1016/j.jbi.2015.11.004] [Medline: 26556646]
47. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Van LH, et al. Evaluation of Facebook and Twitter monitoring to detect safety signals for medical products: An analysis of recent FDA safety alerts. Drug Saf 2017 Apr;40(4):317-331 [FREE Full text] [doi: 10.1007/s40264-016-0491-0] [Medline: 28044249]
48. Bian J, Topaloglu U, Yu F. Towards Large-scale Twitter mining for drug-related adverse events. SHB12 2012 Oct 29;2012:25-32 [FREE Full text] [doi: 10.1145/2389707.2389713] [Medline: 28967001]
49. Leis A, Ronzano F, Mayer MA, Furlong LI, Sanz F. Detecting signs of depression in tweets in Spanish: Behavioral and linguistic analysis. J Med Internet Res 2019 Jun 27;21(6):e14199 [FREE Full text] [doi: 10.2196/14199] [Medline: 31250832]
50. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on twitter? Mining tweets for adverse drug reactions. AMIA Annu Symp Proc 2014;2014:924-393 [FREE Full text] [Medline: 25954400]
51. Mahata D, Friedrichs J, Shah RR, Jiang J. Detecting personal intake of medicine from Twitter. IEEE Intell Syst 2018 Jul;33(4):87-95. [doi: 10.1109/mis.2018.043741326]
52. Klein A, Sarker A, Rouhizadeh M, O'Connor K, Gonzalez G. Detecting personal medication intake in Twitter: an annotated corpusbaseline classification system. 2017 Presented at: the 16th Biomedical Natural Language Processing (BioNLP); Aug 4; Vancouver, BC, Canada p. 1-11 URL: https://www.aclweb.org/anthology/W17-2316.pdf
53. Kiritchenko S, Mohammad S, Morin J, de BB. NRC-Canada at SMM4H shared task: classifying tweets mentioning adverse drug reactions and medication intake. 2017 Presented at: In Proceedings of 2nd Social Media Mining for Health Applications Workshop co-located with the American Medical Informatics Association Annual Symposium (AMIA); Nov 4, 2017; Washington DC p. 1-11 URL: http://ceur-ws.org/Vol-1996/paper1.pdf
54. Nikfarjam A, Ransohoff JD, Callahan A, Jones E, Loew B, Kwong BY, et al. Early detection of adverse drug reactions in social health networks: A natural language processing pipeline for dignal detection. JMIR Public Health Surveill 2019 Jun 03;5(2):e11264 [FREE Full text] [doi: 10.2196/11264] [Medline: 31162134]
55. Segura-Bedmar I, Martínez P, Revert R, Moreno-Schneider J. Exploring Spanish health social media for detecting drug effects. BMC Med Inform Decis Mak 2015;15 Suppl 2:S6 [FREE Full text] [doi: 10.1186/1472-6947-15-S2-S6] [Medline: 26100267]
56. Statista. Leading Countries Based on Number of Twitter Users as of April 2020 (in Millions). 2020. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [accessed 2020-07-05]

57. Liu J, Weitzman E, Chunara R. Assessing behavioral stages from social media data. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW 2017; February 25-March 1, 2017; Portland, Oregon p. 1320-1333.
58. Saha K, Sugar B, Torous J, Abrahao B, Kiciman E, De Choudhury M. A social media study on the effects of psychiatric medication use. In: Proceedings of the thirteenth International Conference on Web and Social Media. 2019 Presented at: ICWSM 2019; 2019; Munich, Germany p. 440-451.
59. Twitter Developer. URL: https://developer.twitter.com/en.html [accessed 2019-12-16]
60. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008;36(Database issue):D901-D906 [FREE Full text] [doi: 10.1093/nar/gkm958] [Medline: 18048412]
61. WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health. The Anatomical Therapeutics Chemical Classification System (ATC). Last updated December 16. 2019 Jan 16. URL: https://www.whocc.no/ [accessed 2019-05-06]
62. Selective serotonin reuptake inhibitor.: Wikipedia; 2019 May 07. URL: https://en.wikipedia.org/wiki/Selective_serotonin_reuptake_inhibitor [accessed 2019-05-07]
63. Spanish Agency for Medicines and Health Products (AEMPS), Ministry of Health and Consumer Affairs and Social Welfare. Multiregional primary care database of Spanish population (BIFAP). 2019. URL: http://www.bifap.org/ [accessed 2019-05-07]
64. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. J R Stat Soc B 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
65. Natural Language Tool Kit. URL: https://www.nltk.org/api/nltk.tokenize.html [accessed 2019-12-10]
66. Padró L, Stanilovsky E. FreeLing 3.0: Towards wider multilinguality. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation. 2012 Presented at: LREC'12; May 21-27; Istanbul, Turkey p. 2473-2479 URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/430_Paper.pdf
67. Perez-Rosas V, Banea C, Mihalcea R. Learning sentiment lexicons in Spanish. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation. 2012 Presented at: LREC'12; May 21-27; Istanbul, Turkey p. 3077-3081 URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/1081_Paper.pdf
68. Ekman P, Friesen WV, O'Sullivan M, Chan A, et al. Universals and cultural differences in the judgments of facial expressions of emotion. J Pers Soc Psychol 1987 Oct;53(4):712-717. [doi: 10.1037//0022-3514.53.4.712] [Medline: 3681648]
69. Sidorov G, Miranda-Jiménez S, Viveros-Jiménez F, Gelbukh A, Castro-Sánchez N, Castillo F. Empirical study of opinion mining in Spanish tweets. In: Proceedings of the 11th Mexican International Conference on Artificial Intelligence. 2012 Presented at: MICAI'12; October 27-November 4, 2012; San Luis Potosí, Mexico p. 1-4. [doi: 10.1007/978-3-642-37798-3_4]
70. Carbonell P, Mayer MA, Bravo A. Exploring brand-name drug mentions on Twitter for pharmacovigilance. Stud Health Technol Inform 2015;210:55-59. [Medline: 25991101]
71. Morris DW, Rush AJ, Jain S, Fava M, Wisniewski SR, Balasubramani GK, et al. Diurnal mood variation in outpatients with major depressive disorder: Implications for DSM-V from an analysis of the sequenced treatment alternatives to relieve depression study data. J Clin Psychiatry 2007 Sep;68(9):1339-1347. [Medline: 17915971]
72. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
73. Anghelescu IG, Kohnen R, Szegedi A, Klement S, Kieser M. Comparison of Hypericum extract WS 5570 and paroxetine in ongoing treatment after recovery from an episode of moderate to severe depression: results from a randomized multicenter study. Pharmacopsychiatry 2006 Nov;39(6):213-219. [doi: 10.1055/s-2006-951388] [Medline: 17124643]
74. Sansone RA, Sansone LA. Antidepressant adherence: are patients taking their medications? Innov Clin Neurosci 2012 May;9(5-6):41-46 [FREE Full text] [Medline: 22808448]
75. Mitchell AJ. Depressed patients and treatment adherence. Lancet 2006 Jun 24;367(9528):2041-2043. [doi: 10.1016/S0140-6736(06)68902-2] [Medline: 16798371]
76. De las Cuevas C, Peñate W, Sanz EJ. Risk factors for non-adherence to antidepressant treatment in patients with mood disorders. Eur J Clin Pharmacol 2014 Jan;70(1):89-98. [doi: 10.1007/s00228-013-1582-9] [Medline: 24013851]
77. De Choudhury M, De S. Mental health discourse on Reddit: Self-disclosure, social support, and anonymity. In: Proceedings of the 8th International Conference on Weblogs and Social Media. 2014 Presented at: AAAI'14; June 1-4; Ann Arbor, Michigan p. 71-80 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8075/8107
78. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. A study of the demographics of web-based health-related social media users. J Med Internet Res 2015 Aug 06;17(8):e194 [FREE Full text] [doi: 10.2196/jmir.4308] [Medline: 26250986]
79. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
80. Yom-Tov E, Johansson-Cox I, Lampos V, Hayward AC. Estimating the secondary attack rate and serial interval of influenza-like illnesses using social media. Influenza Other Respir Viruses 2015 Jul;9(4):191-199 [FREE Full text] [doi: 10.1111/irv.12321] [Medline: 25962320]

81. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]

Abbreviations

POS: part-of-speech
SSRIs: selective serotonin reuptake inhibitors

Edited by G Eysenbach, R Kukafka; submitted 01.06.20; peer-reviewed by F Lopez Segui, E Yom-Tov; comments to author 22.06.20; revised version received 01.09.20; accepted 12.11.20; published 18.12.20.

Please cite as:
Leis A, Ronzano F, Mayer MA, Furlong LI, Sanz F
Evaluating Behavioral and Linguistic Changes During Drug Treatment for Depression Using Tweets in Spanish: Pairwise Comparison Study
J Med Internet Res 2020;22(12):e20920
URL: http://www.jmir.org/2020/12/e20920/
doi:10.2196/20920
PMID:33337338

Original Paper

Rhetorical Appeals and Tactics in New York Times Comments About Vaccines: Qualitative Analysis

John Gallagher[1], PhD; Heidi Y Lawrence[2], PhD

[1]Department of English, University of Illinois, Urbana-Champaign, Urbana, IL, United States

[2]Department of English, George Mason University, Fairfax, VA, United States

Corresponding Author:
John Gallagher, PhD
Department of English, University of Illinois, Urbana-Champaign
608 South Wright Street
Urbana, IL, 61801
United States
Phone: 1 267 640 9142
Email: johng@illinois.edu

Abstract

Background: Improving persuasion in response to vaccine skepticism is a long-standing problem. Elective nonvaccination emerging from skepticism about vaccine safety and efficacy jeopardizes herd immunity, exposing those who are most vulnerable to the risk of serious diseases.

Objective: This article analyzes vaccine sentiments in the New York Times as a way of improving understanding of why existing persuasive approaches may be ineffective and offers insight into how existing methods might be improved. We categorize pro-vaccine and anti-vaccine arguments, offering an in-depth analysis of pro-vaccine appeals and tactics in particular to enhance current understanding of arguments that support vaccines.

Methods: Qualitative thematic analyses were used to analyze themes in rhetorical appeals across 808 vaccine-specific comments. Pro-vaccine and anti-vaccine comments were categorized to provide a broad analysis of the overall context of vaccine comments across viewpoints, with in-depth rhetorical analysis of pro-vaccine comments to address current gaps in understanding of pro-vaccine arguments in particular.

Results: Appeals across 808 anti-vaccine and pro-vaccine comments were similar, though these appeals diverged in tactics and conclusions. Anti-vaccine arguments were more heterogeneous, deploying a wide range of arguments against vaccines. Additional analysis of pro-vaccine comments reveals that these comments use rhetorical strategies that could be counterproductive to producing persuasion. Pro-vaccine comments more frequently used tactics such as ad hominem arguments levied at those who refuse vaccines or used appeals to science to correct beliefs in vaccine skepticism, both of which can be ineffective when attempting to persuade a skeptical audience.

Conclusions: Further study of pro-vaccine argumentation appeals and tactics could illuminate how persuasiveness could be improved in online forums.

(J Med Internet Res 2020;22(12):e19504) doi:10.2196/19504

KEYWORDS

vaccination; qualitative research; quantitative research; rhetoric; online comments; anti-vaccination; pro-vaccination

Introduction

Background

Improving persuasive techniques when communicating with the public about vaccines is a long-standing concern. Although vaccination rates across the United States remain high, pockets of elective nonvaccination remain, which facilitate dangerous outbreaks [1,2]. Although newer movements that expand mandates appear to be successful in some locales [3], there is a continued need for persuasion in communities that resist such mandates and in cases where mandates are less viable, such as in the case of adult vaccinations.

This article asks specific questions related to online communication and vaccines: How do commenters, as readers of online newspaper articles, argue about vaccines? How might those arguments be better attuned to opportunities for persuasion? Finally, how can we better understand pro-vaccine arguments within a subset of both supportive and skeptical comments about vaccines? The study presented here analyzes 808 vaccine-specific comments posted on the *New York Times* (NYT) website's online comment section. This analysis reveals two primary findings: (1) Both pro-vaccine and anti-vaccine comments rely on a thematically similar range of tactics, including appeals to children, science, and risks, though often drawing opposing conclusions. (2) Pro-vaccine comments rely on a set of appeals that offer uneven opportunities for audiences to be engaged and persuaded by pro-vaccine arguments.

As a space where people have active and agonistic arguments, online newspaper comment sections offer important insight into the persuasive contexts of vaccination discourses in real-world settings. Developing knowledge about persuasive tactics online can help researchers develop categories for persuasive appeals that users deploy when they discuss vaccines from pro-vaccine and anti-vaccine perspectives. In this study, vaccine comments were categorized according to pro-vaccine or anti-vaccine sentiments expressed, with common strategies and tactics used across comments identified. Critically, this work also addresses an important gap in online health communication by further documenting and analyzing the rhetorical appeals and tactics of *pro-vaccination* argumentation. While pro-vaccine websites [4], anti-vaccine websites [5,6], and anti-vaccination discussion forums [7-9] have been studied, to date no extensive studies have been conducted about the rhetorical appeals and argumentation strategies of pro-vaccination comments within a context of both pro-vaccine and anti-vaccine comments.

Literature Review

What is known about vaccination sentiments and how they are communicated across a variety of spaces, including on the internet, is largely confined to understanding those who express skepticism about vaccines (eg, anti-vaxxers, vaccine refusers, or elective nonvaccinators). Existing studies locate sources of skepticism about vaccines in a broad range of concerns; the well-known yet refuted concern about the measles-mumps-rubella (MMR) vaccine and autism, "toxic overload" caused by vaccination, appeals to "natural" immunity and forms of disease protection, and personal choice/freedom are some of the primary reasons that parents refuse vaccines [8,10-17]. Other bodies of literature focus on the reasons for low uptake of flu vaccine and the rationales that adult patients have for not accepting the vaccine, including perceived constraints, concerns about a range of side effects, and lack of efficacy [18-20].

Still more articles focus on the deficits of those who refuse or question vaccines, examining "reasoning flaws" associated with vaccine concern [21], arguing for the ethical grounds for mandating vaccines [22,23], and counseling physicians and other healthcare providers on how to respond to vaccine concerns [24,25]. These studies frequently use analyses of hesitancy as a basis for these findings. These studies exist alongside popular press publications articulating the scientific and ethical problems associated with vaccine refusal or an "anti-vaccine movement" [26-28]. Another body of literature has studied specifically how these "anti-vaccination" beliefs are articulated online: typologizing skeptical sentiments that are popularly expressed online [5,6]; examining how risks are weighed and articulated in online chat forums [29]; and looking at ways skeptics use differing media platforms to spread anti-vaccination beliefs [7,30]. These studies offer a broad sense of the arguments and beliefs that parents present as rationales for not vaccinating or at least harboring concerns and skepticism about vaccination.

Newer research in this area has addressed differing tactics for online engagement [4-6,31-33], thereby widening the scope of analysis to both "pro" and "anti" sides of the issue on social media platforms. These studies demonstrate how online communication affords researchers new opportunities for: understanding how people communicate about vaccines; understanding a wider range of vaccine sentiments (outside of negative or skeptical ones only); and identifying new opportunities for persuasion or education about vaccines in online sites for public communication and interaction.

The work reported here contributes to this body of research by examining vaccine sentiments in online spaces, first categorizing pro-vaccine and anti-vaccine arguments and then discussing the ways in which their appeals and tactics interact to offer an understanding of the argumentative context in comments. This research also adds in-depth analyses of appeals and tactics in support of vaccines. Such understanding is important for gaining a broader sense of what the public thinks about the issue of vaccines in formats of online newspaper comments and how persuasive discourse works in online spaces. This article offers possible alternate sources of persuasion when those who support vaccines interact with those who are skeptical.

Methods

Design

The NYT was chosen for its established presence in the United States as a space of debate (it is a "paper of record"). The newspaper has a fourteen-person moderating team as well as machine learning technology from Google called Jigsaw. These tools work to keep comments relatively civil insofar as they filter out explicit language, egregious name-calling, and solicitation posts. We hypothesized that commenters from this newspaper would have a pro-vaccine inclination, a hypothesis that was borne out in our qualitative coding (see Qualitative Analysis Methods section).

In order to develop a targeted data set, we used a larger data set of comments (445,441) made on the NYT from May 1, 2015, to August 31, 2015. Comments were web-scraped using the newspaper's application programming interface directions [34]. Comments were collected during the month of September 2015. Our rationale for this time period was to collect all comments for 4 months, which we believed to be a large enough sample to be representative of larger trends in commenting behavior. We then searched for the words "vaccine" and "vaxx" using a wildcard operator: vaccin*, vaxx*, vax*. This process yielded a final data set of 1101 comments about vaccines from the original 445,441 comments gathered. We have placed these comments into a publicly accessible database [35]. In 2 cases, a pro-vaccine comment was repeated, but as the repeated comments were on different threads, we included them in the data set. To develop a deeper analysis of this data, we conducted a qualitative analysis on the data. We note that these comments, whether pro-vaccine or anti-vaccine, are generally civil, meaning they avoid name-calling and offensive language and are articulated in ways that do not employ heated exchange (capitalizations, use of multiple exclamation points, etc). This civility is likely a product of the NYT's moderation of the forum but could also be indicative of civility concerning this topic in this space during this timeframe.

Qualitative Analysis Methods

We coded the comments as pro-vaccine, anti-vaccine, or not applicable. To classify anti-vaccine comments, we decided comments needed to be against vaccines or demonstrate some skepticism toward them. To classify pro-vaccine comments, we looked for comments that advocated the use of vaccines, broadly conceived. The two authors of this paper initially coded the first 100 comments separately and then met to discuss whether we thought this was a productive schema. We confirmed that we were in general agreement about the schema and then proceeded to code the remaining comments. All coding was done separately. We then compared the coding. We disagreed on 21.80% (240/1101) of comments. We did not code these comments thematically due to rater disagreement. The remainder of the comments, which had been agreed upon, broke down as follows: 602 (54.68%) were pro-vaccine, 206 (18.71%) were anti-vaccine, and 53 (4.81%) were neither pro-vaccine nor anti-vaccine and were not coded thematically.

This process yielded 808 comments that we categorized thematically. After the previous categorization schema, we then coded for themes, appeals, and tactics, drawing on a version of guided grounded theory and specifically looking for rhetorical appeals or arguments that speakers used in their posts. We separately coded the initial 100 comments according to appeals and argumentation. Comments could be coded multiple times to account for multiple appeals. We then met to discuss this coding. We eventually settled on three primary appeals as emerging from the data: *dissoi logoi*, appeals to science, and appeals to the "public good"; these appeals were shared by pro-vaccine and anti-vaccine comments yet were obviously used to achieve different ends or conclusions about vaccines. We subsequently proceeded to thematically code the remaining data set, assigning individual themes for the authors to code. We also made notes of less frequent appeals and tactics as they emerged in the data. Within this context of understanding the appeals and tactics used across vaccine comments, we pursued additional, in-depth analysis of pro-vaccine comments because of their novel abundance in this data set.

Results

Below are the pro-vaccine (Textbox 1) and anti-vaccine (Textbox 2) appeals and tactics. The numbers in parentheses are the numbers of comments expressing each appeal or tactic. The appeals and tactics were not mutually exclusive, and many comments contained multiple appeals and tactics. We do not list appeals or tactics that occurred fewer than 5 times.

Textbox 1. Appeals and tactics of pro-vaccine comments (n=602).

Appeals

- The public good (234)
- Science and expertise advocacy (166)
- *Dissoi logoi* (constructing opposing viewpoints) (102)
- Personal experience or personal ethos (79)
- The social good (68)
- Accepting small risks (50)
- Attributing vaccine denial to the other political spectrum (31)

Tactics

- Direct (use of username or name) or indirect (use of second person) address of vaccine refusers (129)
- Ad hominem (name-calling) (79)
- Debunking autism-vaccine connections (48)
- Referring to evidence (use of hyperlink, typically to government websites) (39)
- Asking anti-vaccine proponents to conduct research (17)
- Debunking mercury-vaccine connections (6)
- Debunking thimerosal connection (6)

Textbox 2. Appeals and tactics of anti-vaccine comments (n=206).

Appeals

- Skepticism: toward institutions (56); toward science (22); toward vaccination rates (14); toward herd immunity (8)
- Concern over money in “big pharma” or “big money” (the money in the pharmacy industry; used either phrase) (33)
- Risk to pregnant individuals (31)
- Personal freedom (29)
- Concern over number of vaccines administered (28)
- Appealing to “complex issues” (some form of the word “complex” without a specific issue mentioned) (22)
- Concern over thimerosal, mercury, and additives in vaccines (17)
- Criticism that vaccines are often for preventable diseases (14)
- Concern over thimerosal in general (12)
- There needs to be more market competition for testing and developing vaccines (11)
- Attempts to connect vaccines to non-autism diseases (7)
- Risk is logical (6)
- Children at risk of vaccines (5)

Tactics

- Claiming not to be 100% against vaccines (31)
- Referring to evidence (use of hyperlink, typically to commercial websites) (20)
- Scientific terminology used incorrectly (17)
- Direct or indirect address to pro-vaccine commenter (use of username or second person) (16)
- Asking other commenters to conduct research (11)
- Against ad hominem (8)
- Ad hominem (7)
- Criticism of “other side” (7)
- Direct address to author of article (use of author’s first name, last name, or both) (5)

Discussion

Qualitative analyses demonstrate that there were common appeals and tactics used across pro-vaccine and anti-vaccine discourses, but also some notable differences in the arguments and argumentative patterns of different positions. This analysis focuses in particular on a number of trends in pro-vaccine argumentation that make opportunities for persuasion problematic and that, if addressed, could help improve the persuasive quality of pro-vaccine discourse online.

Overview

Pro-vaccine comments drew on a narrower range of appeals, tactics, and themes whereas anti-vaccine comments had a broader range of appeals, tactics, and themes. We attribute this difference to at least two causes. First, our binary coding schema of pro-vaccine and anti-vaccine led us to categorize many comments that expressed skepticism over some vaccines or approaches to vaccination as “anti-vaccine.” Comments that were vehemently, completely anti-vaccine were lumped into the same category as comments that were pro-vaccine for childhood vaccinations (eg, MMR, chicken pox) while being anti-vaccine for newer vaccinations (eg, human papillomavirus [HPV]). Moreover, we found many comments that expressed hesitation over vaccination scheduling. Second, being pro-vaccine involves advocating *for something.* Conversely, anti-vaccine comments simply need to cast skepticism on that something. Therefore, the wider set of appeals, tactics, and themes used by anti-vaccination comments makes sense: skepticism involves creating some level of doubt, even if minor. Anti-vaccine comments thus need only discuss a concerning issue, no matter how small or insignificant it is.

In broad strokes, anti-vaccine comment strategies are more diverse and heterogeneous than pro-vaccine comment strategies. Anti-vaccine commenters appealed to science’s fallibility, the minute presence of risk, and freedom of choice. Anti-vaccine commenters appealed to the debunked research conducted by Andrew Wakefield and others. Anti-vaccine commenters used hyperbole frequently, notably in increasing the number and frequency of vaccine shots. The concept of risk is particularly illustrative. On the one hand, pro-vaccine comments discuss risk in a relatively homogeneous way: it is something to be minimized, and comments recognize the low level of risk. On the other hand, anti-vaccine comments were concerned over realistic types of risk *and* radically untrue types of risk, including overstatement and hyperbole, which was dramatically uneven in the data set. Some anti-vaccine commenters argued that physicians wanted to give babies dozens of vaccines and argued that all vaccines had high risks. Other commenters expressed less hyperbolic concerns, such as worries about spacing of the MMR vaccine and a desire to discuss risk with a practitioner before accepting a vaccine.

Pro-Vaccine Comments

We noted early in our analysis that the data set included a novel quantity of pro-vaccine comments. We suspect this number of pro-vaccine comments is a product of the NYT's typical readership as well as the effect of comment moderation. However, as noted in the literature review above, with the typical focus on anti-vaccine comments and deficit approaches to correcting this stance, less information is established in the literature on the argumentative patterns of pro-vaccine advocates. We report on these trends below to add novel findings to the literature, but also to point to spaces where pro-vaccine advocates could improve the persuasiveness of their commentary.

Overall, we found that pro-vaccine comments included three dominant appeals and tactics. The first category was appeals to science and expertise. Science and expertise advocacy includes appeals to science as a black box (eg, advocating for trusting "the science", as in the words themselves). These appeals also include trusting scientists and experts in general (most frequently the "medical community") as well as specifically by a proper name. These appeals included advocating for rationality and logic and attacking anti-science viewpoints (eg, "science-deniers", lack of "scientific links", "anti-science ignorance"). Less frequently, commenters in this category mentioned scientific concepts such as controlled studies, falsifiability, and hypothesis testing. Comments that used this appeal were heavily anti-anecdote and requested evidence. They corrected scientific inaccuracies (such as the number of required vaccines, side effects of vaccines, and use of mercury). We note that this operates in opposition to skeptical comments about trusting individual forms of knowledge, such as personal experience. This creates a clash between these two perspectives, wherein the concept of science becomes a stand-in for trustworthy forms of expertise, and skeptical perspectives denounce or diminish that perspective through—often incorrect—critiques of science. Although such corrective forms of communication might seem like a helpful, even persuasive, intervention, previous studies of vaccine sentiment have indicated that such measures can have a "backfire effect," causing people to more firmly believe incorrect beliefs upon having them corrected [36].

Second, for vaccine proponents, vaccination is associated with the common good or what is generally perceived to be good for the public. A remarkable number of pro-vaccine comments focused on appealing to the public and social good of vaccines. *Public good* took multiple forms, including herd immunity, the safety of children and older adults, and the sickness and deaths of young children from vaccine-preventable diseases. These comments often mentioned legal liability for those who do not vaccinate; the extension of this logic implies a prevention of epidemics and increase in herd immunity. Comments that appealed to the public good of vaccines frequently made other non–vaccine-related claims about the public good, including climate change and economic equality.

Comments that focused on what we label the *social good*, while related to the public good, made use of historical statistics and information to discuss the quality of life that vaccines brought about. We note that this appeal is often used in conjunction with personal experience and *ethos*. Commenters frequently recall their childhoods, often detailing the suffering of other children from vaccine-preventable diseases. They do so in order to discuss the social good (and social progress) that vaccines have society. These appeals, though they offer the potential for persuasion through their use of anecdote and narrative—devices that can be more persuasive for a skeptical public—also operate in direct opposition to skeptical appeals to individualism rather than collective good. Thus, the ideological gap between the "good story" these conflicting narratives tell (one of protecting individual rights versus achieving collective good) lessens their potential to operate persuasively for skeptical readers.

Third, vaccine proponents construct opponents' arguments as a way of establishing their position and amplifying their support of vaccines, an appeal called *dissoi logoi*, or construction of oppositional or contrasting arguments [37]. *Dissoi logoi* in classical argumentation originated as a mechanism for understanding and examining opposing sides. In its best form, *dissoi logoi* allows the speaker to see and articulate an issue from someone else's perspective, but it can lead to specious arguments as well when oppositional arguments are misunderstood, weakly constructed, or incorporate their own fallacies. When pro-vaccine arguments employed *dissoi logoi*, they used it to approximate or describe vaccine skepticism from within their own position of support, leading to reductive and ad hominem attacks associated with their estimation of the "opposing side." These appeals are particularly potent sources of creating opposition, rather than opportunities for persuasion, since restated arguments frequently create opposing arguments not worthy of refutation.

For example, the following comment conflates concern about HPV vaccine with laziness or ignorance:

> *[Rick Perry, Governor of Texas who endorsed an unpopular mandate for HPV vaccine,] should have gone through with it for their own good. This vaccine, if taken early enough, will prevent cancer. The people in Texas who are opposed are a bunch of religious bozos who think that getting this vaccine will make it more likely that their children will have premarital sex. That is sheer ignorance.*

This comment engages in *dissoi logoi* through constructing the argument of vaccine skeptics, invoking "the people in Texas who are opposed" to the policy. This comment features multiple ad hominem tactics insulting anti-vaccine perspectives or anti-vaccine commenters, including calling skeptics "bozos" and citing them as a danger to efforts to prevent cancer. Ad hominem was significantly more prevalent in pro-vaccine comments than in anti-vaccine comments and is classically problematic as a persuasive act, since personal attacks can cause defensiveness, thus diminishing persuasive appeal.

Finally, it is worth noting that there also appear to be ideological components to both the pro-vaccine and anti-vaccine comments. The ideology that undergirds pro-vaccine comments includes concern for public welfare and society as a functioning whole, whereas the anti-vaccine comments have an ideology of concern for individual welfare and autonomy. From our reading of the comments, these ideologies operate independently of typical US politics (conservative vs liberal) because many comments attribute the anti-vaccine perspective to the "other side." Often, if a comment appears to have a traditionally conservative leaning, then it will attribute the anti-vaccine movement to liberals, noting that liberal places in California or Oregon have outbreaks of preventable diseases. Anti-vaccine perspectives were also attributed by pro-vaccine commenters to people who are anti-science, anti–genetically modified organisms (GMOs), and anti–nuclear power, and climate change deniers. For instance, one emblematic comment reads as follows:

> *I find it very hard to respect the resistance to empiricism common to the anti-GMO crowd, anti-vaccine fanatics and climate change deniers. The widespread popular refusal to come to conclusions based on evidence suggests a dim future for us all. We will need science, we will need technology, and GMOs will have to be one of the tools available to us if we hope to feed the nine billion people expected to share a warming and highly stressed planet at mid-century.*

On the other hand, if a comment appears to have a liberal leaning, it will attribute the anti-vaccine movement to conservatives. These comments often noted that the concept of "personal choice" undermines herd immunity and the health of the public. As an emblematic example, one commenter writes the following:

> *Well, in large part because of the policies--family planning, vaccination, the Peace Corps, the green semi-revolution and so on--that the Right foight [sic] tooth and nail against spending any money on? By the way: that progress, and in many ways it IS [sic] progress, did come at a price, you know. Take a look at the bill, which includes warmong [sic] up the planet and polluting a fair old chunk of it.*

We make this observation as a way of noting that, although varying ideologies are operating within each perspective, a clear political alignment cannot be discerned.

Limitations and Future Research

While there are several limitations of this study, we first address a strength of this study: the agreement between the two authors. We had not worked together before but came to surprising agreement (861/1101 comments, 78.20%) about whether to categorize comments as pro-vaccine or anti-vaccine (or not applicable). Following this degree of agreement, we believe that most vaccine-related comments in our data set ascribe to either advocacy of vaccines or skepticism of vaccines.

We urge caution with our findings due to the limited size of the data set (808 comments), the venue in which the comments were made (the NYT), and our own coding schema that did not analyze comments that we disagreed upon or comments that we categorized as neither pro-vaccine nor anti-vaccine. The greater number of pro-vaccine comments may be the result of the screening process by the moderating team at the NYT, the Jigsaw machine learning technology, or both. More conceptually, however, our method has not addressed the many other issues present in these comments, including US politics, local discussions, reader-to-reader relationships, and so forth. In this sense, we have focused our analysis to a particular theme at the detriment of examining how those themes intersect with other rhetorical moves made.

A more conceptual limitation is that, because the data we used were public comments, the commenters may not see themselves as attempting to be truly rhetorical (eg, persuasive). In this sense, these findings have qualifications, the most relevant being that the appeals and tactics are not necessarily geared at being highly persuasive. Nevertheless, the detail offered in these comments adds information to the online conversation around vaccination, notably the finding that the pro-vaccine and anti-vaccine comments deploy several similar strategies while providing foils to one another.

A major challenge about this project has to do with the heterogeneous debates under the broad term of vaccine. Within our data set, vaccines were discussed, many of which were childhood vaccinations (MMR, chickenpox, etc). However, other types of vaccines were discussed, such as the Ebola vaccine (the NYT ran an article about this in 2015). Further, vaccines were mentioned in tandem topics that were not directly related (for instance, vaccines were frequently discussed alongside genetically modified food topics). In our methodology, we have flattened the discussion focus in order to analyze the appeals and tactics surrounding vaccine debate.

As a result of these limitations, we have several suggestions for future research. These include increasing the number of comments analyzed, examining other venues, and running more advanced computer analysis (corpus linguistics and natural language processing). Analyzing vaccine-related comments outside the binary of pro-vaccine and anti-vaccine may be useful for future research. In this article, we have also not analyzed the comments about which we disagreed in our coding schema or the comments that we agreed were neither pro-vaccine nor anti-vaccine. Analyzing these types of comments may reveal additional rhetorical appeals or analytical tactics as well as nuanced relationships between the binary of pro-vaccine and anti-vaccine perspectives. Finally, because comments are an informal place where persuasion might happen, researchers could insert different comments that deploy varying levels of persuasive appeals in order to determine which appeals work more effectively than others. Research could then document reactions to these comments as well as survey and interview commenters to evaluate the effectiveness of specific appeals.

Conclusion

When it comes to vaccination, the stakes are high. With the urgent public health need to achieve high rates of universal vaccination, each encounter with a person about vaccines can be an opportunity to strengthen public trust in and acceptance of vaccines. We offer the previous analyses as a way for pro-vaccine commenters across public and professional spaces to consider how their tactics can be more persuasive to skeptics. Simply dismissing skeptics will not change their minds, and appealing to vaccine supporters does not actively engage with changing the status quo. Our analyses of the pro-vaccine comments may help to guide future studies attempting to identify how and why vaccine skeptics are persuaded in online, participatory environments.

Conflicts of Interest

None declared.

References

1. Omer SB, Salmon DA, Orenstein WA, deHart MP, Halsey N. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. N Engl J Med 2009 May 7;360(19):1981-1988. [doi: 10.1056/NEJMsa0806477] [Medline: 19420367]
2. Hill HA, Elam-Evans LD, Yankey D, Singleton JA, Kang Y. Vaccination Coverage Among Children Aged 19-35 Months - United States, 2017. MMWR Morb Mortal Wkly Rep 2018 Oct 12;67(40):1123-1128 [FREE Full text] [doi: 10.15585/mmwr.mm6740a4] [Medline: 30307907]
3. Shaw J, Mader EM, Bennett BE, Vernyi-Kellogg OK, Yang YT, Morley CP. Immunization Mandates, Vaccination Coverage, and Exemption Rates in the United States. Open Forum Infect Dis 2018 Jun;5(6):ofy130 [FREE Full text] [doi: 10.1093/ofid/ofy130] [Medline: 29977973]
4. Grant L, Hausman BL, Cashion M, Lucchesi N, Patel K, Roberts J. Vaccination persuasion online: a qualitative study of two provaccine and two vaccine-skeptical websites. J Med Internet Res 2015 May 29;17(5):e133 [FREE Full text] [doi: 10.2196/jmir.4153] [Medline: 26024907]
5. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. Vaccine 2010 Feb 17;28(7):1709-1716. [doi: 10.1016/j.vaccine.2009.12.022] [Medline: 20045099]
6. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm--an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012 May 28;30(25):3778-3789. [doi: 10.1016/j.vaccine.2011.11.112] [Medline: 22172504]
7. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The Measles Vaccination Narrative in Twitter: A Quantitative Analysis. JMIR Public Health Surveill 2016;2(1):e1 [FREE Full text] [doi: 10.2196/publichealth.5059] [Medline: 27227144]
8. Smith N, Graham T. Mapping the anti-vaccination movement on Facebook. Inf Commun Soc 2017 Dec 27;22(9):1310-1327 [FREE Full text] [doi: 10.1080/1369118X.2017.1418406]
9. Lei Y, Pereira JA, Quach S, Bettinger JA, Kwong JC, Corace K, et al. Examining Perceptions about Mandatory Influenza Vaccination of Healthcare Workers through Online Comments on News Stories. PLoS ONE 2015 Jun 18;10(6):e0129993. [doi: 10.1371/journal.pone.0129993] [Medline: 26086194]
10. Bocquier A, Ward J, Raude J, Peretti-Watel P, Verger P. Socioeconomic differences in childhood vaccination in developed countries: a systematic review of quantitative studies. Expert Rev Vaccines 2017 Sep 21;16(11):1107-1118. [doi: 10.1080/14760584.2017.1381020] [Medline: 28914112]
11. Largent M. Vaccine: The Debate in Modern America. Baltimore: Johns Hopkins University Press; 2012.
12. Reich J. Calling the Shots: Why Parents Reject Vaccines. New York: New York University Press; 2018.
13. Destefano F, Bodenstab H, Offit P. Principal Controversies in Vaccine Safety in the United States. Clin Infect Dis 2019;69(4):726-731. [doi: 10.1093/cid/ciz135] [Medline: 30753348]
14. Bowes J. Measles, misinformation, and risk: personal belief exemptions and the MMR vaccine. J Law Biosci 2016 Nov 22;3(3):718-725. [doi: 10.1093/jlb/lsw057] [Medline: 28798865]
15. Greenberg J, Dubé E, Driedger M. Vaccine Hesitancy: In Search of the Risk Communication Comfort Zone. PLoS Curr 2017 Mar 03;9:ecurrents.outbreaks.0561a011117a1d1f9596e24949e8690b [FREE Full text] [doi: 10.1371/currents.outbreaks.0561a011117a1d1f9596e24949e8690b] [Medline: 28357154]
16. Poland GA. MMR Vaccine and Autism: Vaccine Nihilism and Postmodern Science. Mayo Clin Proc 2011 Sep;86(9):869-871. [doi: 10.4065/mcp.2011.0467] [Medline: 21878599]
17. Chang LV. Information, education, and health behaviors: Evidence from the MMR vaccine autism controversy. Health Econ 2018 May 02;27(7):1043-1062. [doi: 10.1002/hec.3645] [Medline: 29717799]
18. Maisa A, Milligan S, Quinn A, Boulter D, Johnston J, Treanor C, et al. Vaccination against pertussis and influenza in pregnancy: a qualitative study of barriers and facilitators. Public Health 2018 Sep;162:111-117. [doi: 10.1016/j.puhe.2018.05.025] [Medline: 30007172]
19. Luz PM, Johnson RE, Brown HE. Workplace availability, risk group and perceived barriers predictive of 2016-17 influenza vaccine uptake in the United States: A cross-sectional study. Vaccine 2017 Oct 13;35(43):5890-5896. [doi: 10.1016/j.vaccine.2017.08.078] [Medline: 28890192]
20. Lawrence HY. Vaccine Rhetorics. Ohio: Ohio State University Press; Jan 01, 2020.

21. Jacobson RM, Targonski PV, Poland GA. A taxonomy of reasoning flaws in the anti-vaccine movement. Vaccine 2007 Apr 20;25(16):3146-3152. [doi: 10.1016/j.vaccine.2007.01.046] [Medline: 17292515]
22. Bayefsky MJ. The Ethical Case for Mandating HPV Vaccination. J Law Med Ethics 2018 Jul 17;46(2):501-510. [doi: 10.1177/1073110518782957] [Medline: 30147003]
23. Hendrix KS, Sturm LA, Zimet GD, Meslin EM. Ethics and Childhood Vaccination Policy in the United States. Am J Public Health 2016 Feb;106(2):273-278. [doi: 10.2105/ajph.2015.302952] [Medline: 26691123]
24. Shelby A, Ernst K. Story and science: how providers and parents can utilize storytelling to combat anti-vaccine misinformation. Hum Vaccin Immunother 2014 Oct 27;9(8):1795-1801. [doi: 10.4161/hv.24828] [Medline: 23811786]
25. Logan J, Nederhoff D, Koch B, Griffith B, Wolfson J, Awan FA, et al. 'What have you HEARD about the HERD?' Does education about local influenza vaccination coverage and herd immunity affect willingness to vaccinate? Vaccine 2018 Jun 27;36(28):4118-4125 [FREE Full text] [doi: 10.1016/j.vaccine.2018.05.037] [Medline: 29789242]
26. Offit P. Deadly Choices: How the Anti-Vaccine Movement Threatens Us All. New York: Basic Books; 2011.
27. Mnookin S. The Panic Virus: A True Story of Medicine, Science, and Fear. New York: Simon & Schuster; 2011.
28. Biss E. On Immunity: An Inoculation. Minneapolis, MN: Greywolf Press; 2015.
29. Skea ZC, Entwistle VA, Watt I, Russell E. 'Avoiding harm to others' considerations in relation to parental measles, mumps and rubella (MMR) vaccination discussions - an analysis of an online chat forum. Soc Sci Med 2008 Nov;67(9):1382-1390. [doi: 10.1016/j.socscimed.2008.07.006] [Medline: 18703263]
30. Ache KA, Wallace LS. Human papillomavirus vaccination coverage on YouTube. Am J Prev Med 2008 Oct;35(4):389-392. [doi: 10.1016/j.amepre.2008.06.029] [Medline: 18675530]
31. Love B, Himelboim I, Holton A, Stewart K. Twitter as a source of vaccination information: content drivers and what they are saying. Am J Infect Control 2013 Jun;41(6):568-570. [doi: 10.1016/j.ajic.2012.10.016] [Medline: 23726548]
32. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. J Med Internet Res 2016;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
33. Tustin JL, Crowcroft NS, Gesink D, Johnson I, Keelan J, Lachapelle B. User-Driven Comments on a Facebook Advertisement Recruiting Canadian Parents in a Study on Immunization: Content Analysis. JMIR Public Health Surveill 2018 Sep 20;4(3):e10090 [FREE Full text] [doi: 10.2196/10090] [Medline: 30249585]
34. The New York Times Developer Network. URL: https://developer.nytimes.com [accessed 2015-09-16]
35. Gallagher J. NYT vaccine comments. Urbana-Champaign: University of Illinois at Urbana-Champaign URL: https://databank.illinois.edu/datasets/IDB-7724021 [accessed 2019-09-06]
36. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: a randomized trial. Pediatrics 2014 Apr;133(4):e835-e842 [FREE Full text] [doi: 10.1542/peds.2013-2365] [Medline: 24590751]
37. Anon. Dissoi logoi. In: The rhetorical tradition: Readings from classical times to the present. Boston: Bedford-St.Martin's; 2001:47-55.

Abbreviations

GMO: genetically modified organism
HPV: human papillomavirus
MMR: measles-mumps-rubella
NYT: New York Times

Edited by G Eysenbach; submitted 26.04.20; peer-reviewed by D Surian, L Grant; comments to author 12.06.20; revised version received 04.08.20; accepted 24.10.20; published 04.12.20.

Please cite as:
Gallagher J, Lawrence HY
Rhetorical Appeals and Tactics in New York Times Comments About Vaccines: Qualitative Analysis
J Med Internet Res 2020;22(12):e19504
URL: http://www.jmir.org/2020/12/e19504/
doi:10.2196/19504
PMID:33275110

Original Paper

COVID-19 Misinformation Spread in Eight Countries: Exponential Growth Modeling Study

Elaine Okanyene Nsoesie[1], MS, PhD; Nina Cesare[2], PhD; Martin Müller[3], MSc; Al Ozonoff[4], PhD, CPPS

[1]Department of Global Health, School of Public Health, Boston University, Boston, MA, United States
[2]Biostatistics and Epidemiology Data Analytics Center, School of Public Health, Boston University, Boston, MA, United States
[3]Digital Epidemiology Lab, École Polytechnique Fédérale, Geneva, Switzerland
[4]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Elaine Okanyene Nsoesie, MS, PhD
Department of Global Health
School of Public Health
Boston University
801 Massachusetts Ave
Third floor
Boston, MA, 02118
United States
Phone: 1 617 638 5234
Email: onelaine@bu.edu

Abstract

Background: The epidemic of misinformation about COVID-19 transmission, prevention, and treatment has been going on since the start of the pandemic. However, data on the exposure and impact of misinformation is not readily available.

Objective: We aim to characterize and compare the start, peak, and doubling time of COVID-19 misinformation topics across 8 countries using an exponential growth model usually employed to study infectious disease epidemics.

Methods: COVID-19 misinformation topics were selected from the World Health Organization Mythbusters website. Data representing exposure was obtained from the Google Trends application programming interface for 8 English-speaking countries. Exponential growth models were used in modeling trends for each country.

Results: Searches for “coronavirus AND 5G” started at different times but peaked in the same week for 6 countries. Searches for 5G also had the shortest doubling time across all misinformation topics, with the shortest time in Nigeria and South Africa (approximately 4-5 days). Searches for “coronavirus AND ginger” started at the same time (the week of January 19, 2020) for several countries, but peaks were incongruent, and searches did not always grow exponentially after the initial week. Searches for “coronavirus AND sun” had different start times across countries but peaked at the same time for multiple countries.

Conclusions: Patterns in the start, peak, and doubling time for “coronavirus AND 5G” were different from the other misinformation topics and were mostly consistent across countries assessed, which might be attributable to a lack of public understanding of 5G technology. Understanding the spread of misinformation, similarities and differences across different contexts can help in the development of appropriate interventions for limiting its impact similar to how we address infectious disease epidemics. Furthermore, the rapid proliferation of misinformation that discourages adherence to public health interventions could be predictive of future increases in disease cases.

(J Med Internet Res 2020;22(12):e24425) doi:10.2196/24425

KEYWORDS

misinformation; internet; COVID-19; social media; rumors

Introduction

SARS-CoV-2 has infected more than 18.4 million people worldwide and has resulted in approximately 692,000 deaths [1]. Fast-paced research intended to understand the disease biology and dynamics, the novelty of the pandemic experience, and the quickly evolving physical distancing protocols have meant rapid changes in the public’s understanding of the disease, which has created an environment primed for the spread of misinformation. These include unsubstantiated or false claims that typically relate to one of four topics: transmission, prevention, vaccination, and treatment [2]. For example, there have been claims that COVID-19 was originally developed as a bioweapon [3] and false information about preventive substances or remedies including vitamin C and D, zinc, elderberry, chlorine dioxide, silver, and essential oils [4,5]. The need for information; anxiety about physical, social, and economic impacts of the virus; and a lack of a centralized authority available to detect and combat misinformation created an environment in which false assertions about COVID-19 could spread unchecked [6,7]. Although institutional efforts have been made to combat false claims about COVID-19 through channels such as the World Health Organization (WHO) Mythbusters website [8] and the Food and Drug Administration's Health Fraud Press Announcements [5], the control of misinformation remains a challenge [2].

Although tracking the origins and spread of false beliefs surrounding COVID-19 remains difficult, infodemiology [9,10] may provide a framework for tracking and analyzing social determinants of COVID-19 misinformation spread [11-13]. Moreover, there is evidence that trends in how misinformation spreads online parallels the spread of epidemics [14,15]. We therefore aim to understand how misinformation exposure differs across countries, what similarities and differences exist, and what types of misinformation spread fastest. We used epidemic modeling techniques to characterize misinformation about COVID-19 in 8 countries, focusing on the start, peak, and the doubling time of searches. Characterizing how misinformation-seeking trends develop online can be useful in the design of appropriate interventions that aid in the control of epidemics.

Methods

Data

We constructed a term list consisting of a combination of "Coronavirus," "COVID-19," "COVID19," and "COVID" with misinformation terms obtained from the WHO Mythbusters website [8]: wine, hot weather, antibiotics, chlorine, garlic, ginger, sun, 5G, hydroxychloroquine, pepper, houseflies, mosquito, hand dryer, supplement, and saline. We selected topics that were clearly defined and for which there was available data. For example, someone searching for "COVID AND Alcohol" might be interested in the amount of alcohol needed to make hand sanitizers at home and not necessarily trying to verify the claim that drinking alcohol might cure or prevent COVID-19. Topics such as 5G are reported to have spread quickly through social networks on Twitter [16]. Additionally, searches for terms such as mosquito and hand dryers were prevalent in some countries and completely absent in others.

After assessing the quality of search data across the countries, we focused on four misinformation topics: claims that (1) drinking alcohol (specifically, wine) increases immunity to COVID-19; (2) sun exposure prevents spread or that COVID-19 is less likely to spread in hot, sunny areas; (3) home remedies may prevent or cure COVID-19; and (4) COVID-19 is spread via 5G cellular networks. We also discussed searches for hydroxychloroquine separately because, unlike the other misinformation terms, it did not appear on the WHO website until July 31, 2020. There was also much confusion about its potential benefit while it was being evaluated by clinicians, unlike the other misinformation topics.

We focused on 8 English-speaking countries from five continents: Nigeria, Kenya, South Africa, the United States, the United Kingdom, India, Australia, and Canada. Weekly search data was obtained from the Google Trends application programming interface from December 2019 to October 2020.

Analysis

We assumed the search data represented trends in exposure to misinformation. This implies that if someone is seeking information on a particular misinformation topic, they have been exposed to it. However, we cannot deduce a person's intent or whether or not they believe the misinformation. Additional data is needed to deduce the personal motives of individuals engaging in searches (see Discussion section). We inferred the week of the first peak and then fitted an exponential growth model to both sides of the time series curve: before and after the peak. The exponential regression model is defined as follows: $\log(y) = r \times t \times b$, where y represents searches (or postings) for the misinformed phrase, and r, t, and b represent the growth rate, number of days since Google reported a search volume greater than one, and the intercept, respectively. This approach was implemented in the Incidence package in the R software (R Foundation for Statistical Computing) and used for analyzing incidence data for epidemics [17,18]. We compared the start week, doubling time, and first peak across the 8 countries and four topics. We referred to the peak of the search data as the initial peak, since similar to an epidemic, there can be multiple peaks. New searches might be initiated at a later time during the pandemic.

Results

Start and Peak Weeks

Searches for "coronavirus AND 5G" started at different times but peaked in the same week for 6 of the countries (Figure 1). For example, searches for 5G in Australia, the United Kingdom, and Canada were first reported during the week of January 19, 2020, while searches in South Africa, India, and the United States started the following week. In contrast, searches were recorded from Kenya and Nigeria a month later, during the weeks of February 16 and 23, respectively. Despite the different start dates, India, Australia, Canada, Kenya, Nigeria, and the United States observed the first search peak during the week of April 5, 2020. The United Kingdom and South Africa observed a peak during the same week: March 29, 2020.

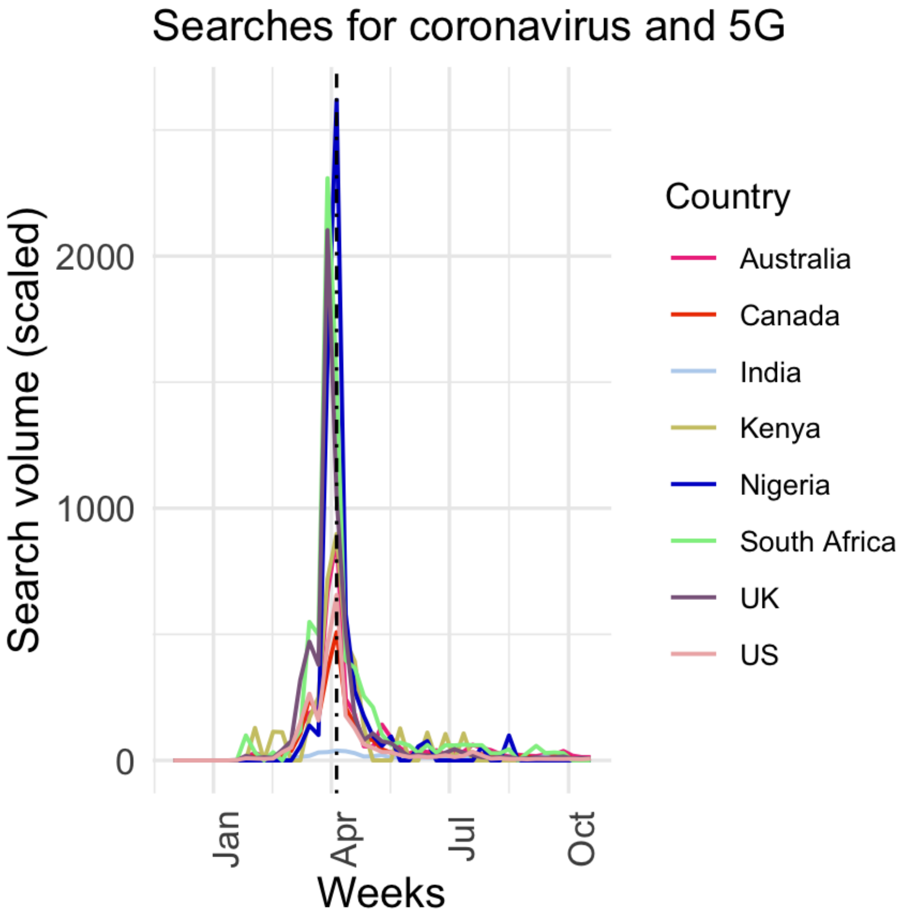
Similarly, searches for "coronavirus AND ginger" started in the same week in several countries. The United States, the United Kingdom, Canada, Australia, and India noted initial searches during the week of January 19, 2020. However, initial searches in South Africa, Nigeria, and Kenya occurred several weeks later, during the weeks of February 9, February 23, and March 8, respectively. The peak week was earliest for the United

Kingdom (March 22), followed by Canada (March 29) and the United States and Australia (April 5). South Africa, Nigeria, India, and Kenya peaked during the weeks of April 12, April 12, April 19, and April 26, respectively. Searches did not always grow exponentially after the initial week. For some countries such as Nigeria, zeros were noted during 1 or 2 consecutive weeks after the initial search. This might be due to Google's scaling algorithm and might not represent no searches during those weeks.

Furthermore, searches for "coronavirus AND sun" started in the United States, Canada, Australia, India, and the United Kingdom during the week of January 19, 2020, in Nigeria and South Africa during the week of January 26, and a month later in Kenya (February 23). Multiple countries noted a peak during the same week: the week of March 15 for the United States, South Africa, and Canada; March 22 for Australia, the United Kingdom, and Nigeria; and April 12 for India and Kenya.

Lastly, search trends for "coronavirus AND wine" were inconsistent across the 8 countries. Nigeria and Kenya had low search volume and were therefore excluded. The United States noted the earliest searches during the week of January 12, 2020. Other countries noted initial search during the weeks of January 19 (Canada, India, the United Kingdom), January 26 (Australia), and February 9 (South Africa). The peak weeks also differed across the countries with no obvious groupings across regions. The peak weeks were March 15 (Canada and South Africa), March 22 (the United Kingdom), April 5 (Australia and the United States), and April 12 (India).

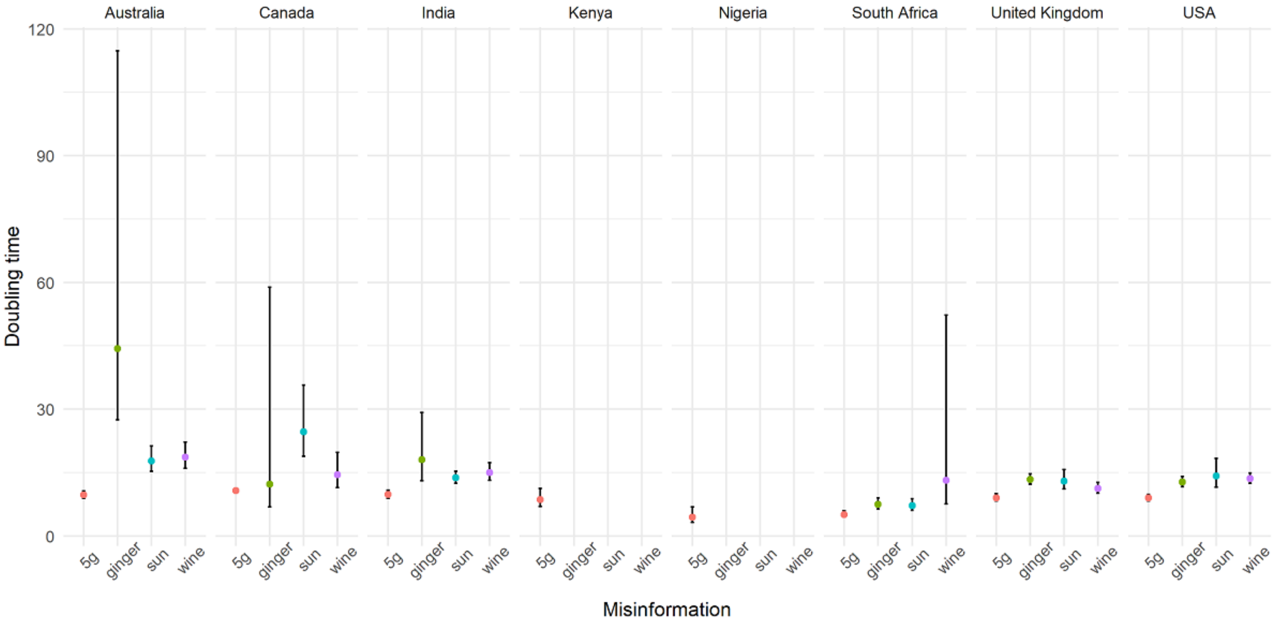
Figure 1. Trends in searches for coronavirus and 5G across 8 countries from December 2019 to October 2020. The black vertical line indicates the time the World Health Organization included the topic on the Mythbusters website.



Doubling Time

Searches for 5G had the shortest doubling time across all misinformation topics (Figure 2). Nigeria and South Africa noted the shortest doubling time: between 4 and 5 days. Searches for ginger doubled at approximately the same rate for the United Kingdom and the United States. Searches for sun doubled much more slowly in Canada compared to the other countries. The confidence interval was also wider, suggesting sparse searches in these contexts and less confidence in the estimates. Similar observations were noted for searches of ginger in Australia. Searches for wine were more prevalent in the United Kingdom, the United States, India, and Australia. The data for Kenya, Nigeria, Canada, and South Africa were sparse.

Figure 2. The estimated doubling time for searches on four misinformation topics: 5G, wine, ginger, and sun.



Discussion

Despite access to the same social media and search platforms, exposure to misinformation appeared different across the 8 countries included in our study. Searches for the majority of misinformation topics varied in start and peak time, and did not necessarily grow exponentially.

These differences in the timing of initial searches could be due to disparities in access, culture, and how internet platforms are used in different parts of the world. Moreover, the sparseness of searches for topics such as ginger in some contexts indicates contextual differences in the concerns and interests.

Searches for 5G and hydroxychloroquine displayed unique patterns that cut across contexts. Not only did searches for 5G have the fastest doubling time, but they also started and peaked around the same time for most of the countries. Searches for hydroxychloroquine had a distinct trend when compared to the other topics (Figure 3), owing to the public and medical discussion of its potential benefits over several months. The WHO listed hydroxychloroquine on their website on July 31, 2020, during what appears to be the second or third wave of searches.

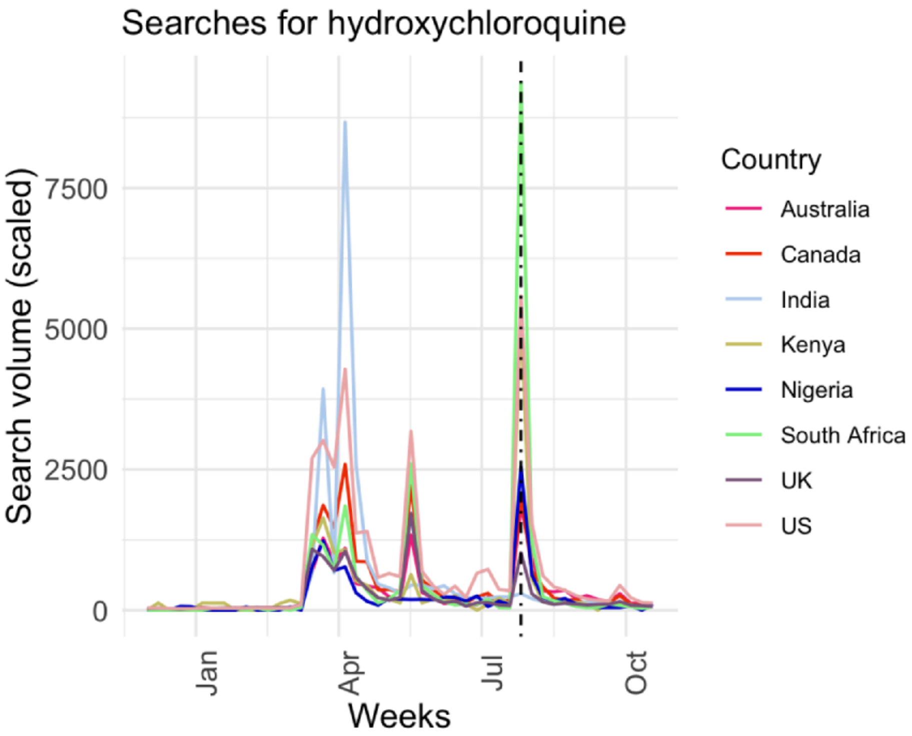
Misinformation about COVID-19 transmission, prevention, and treatment can impact how the public reacts to public health interventions such as wearing a mask and social distancing, which can lead to an uptick in reported cases. Although it is beyond the scope of this paper to link specific misinformation topics to the spread of COVID-19 in different countries, we observed that, for some misinformation topics such as 5G, the official date of the WHO's response noted on the Mythbusters website appeared after the first peak in searches for some countries (Figure 1). This was similar for the other misinformation topics. In addition, we did not observe a resurgence in 5G searches after the first peak. This observation supports the idea that timely response from trusted public health sources are needed to counter the spread of misinformation and that digital platforms may be useful tools that the WHO and other organizations may combat such falsehoods.

Research that combines data from multiple digital platforms is needed to comprehensively study the emergence of various misinformation topics and their association with reports of increased COVID-19 activity in various regions. These studies must be mindful of context, however, given differences in testing capacity, case, and mortality reporting across countries. Nonetheless, such studies could improve our understanding of the potential impact of misinformation on the transmission of SARS-CoV-2. Additionally, future studies may gather behavioral data to assess how the rapid proliferation of misinformation discourages adherence to public health interventions related to COVID-19. These studies must also take into account the culture, policies, and regional variation of each country analyzed.

We acknowledge limitations in our data. First, only people who have access and can afford using the internet are likely to spend time investigating these misinformation topics. This therefore leaves out a large percentage of the population. The percentage of the population within each country that used the internet in 2017 varied significantly among study contexts, ranging from more than 90% in the United Kingdom to less than 10% in Nigeria [19]. Second, the keyword selection and phrases that characterized our data collection may have unintentionally omitted relevant content or included noise. Third, an analysis of the network characteristics of individuals involved in spreading misinformation would better inform intervention strategies.

Identifying where misinformation trends emerge and how quickly they spread can be used to direct crisis communication and provide more effective health care [10]. This study illustrates that neighboring countries can have different misinformation experiences related to similar topics, which can impact control of COVID-19 in these countries. Although monitoring misinformation-seeking behavior via Google Trends is one pathway for identifying belief prevalence and trends, we should monitor information flow across multiple platforms including social media sites such as Facebook, Twitter, and Instagram, and messaging apps such as WhatsApp.

Figure 3. Trends in searches for hydroxychloroquine across 8 countries from December 2019 to October 2020. The black vertical line indicates the time the World Health Organization included the topic on the Mythbusters website.



Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19): situation report – 133. World Health Organization. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200601-covid-19-sitrep-133.pdf?sfvrsn=9a56f2ac_4 [accessed 2020-06-10]
2. Nsoesie EO, Oladeji O. Identifying patterns to prevent the spread of misinformation during epidemics. HKS Misinformation Rev 2020 Apr 27:1. [doi: 10.37016/mr-2020-014]
3. Meyers SL. China spins tale that the U.S. Army started the coronavirus epidemic. The New York Times. 2020 Mar 13. URL: https://www.nytimes.com/2020/03/13/world/asia/coronavirus-china-conspiracy-theory.html [accessed 2020-05-31]
4. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
5. Press announcements - health fraud. US Food and Drug Administration. URL: https://www.fda.gov/consumers/health-fraud-scams/press-announcements-health-fraud [accessed 2020-06-11]
6. Jo W, Lee J, Park J, Kim Y. Online information exchange and anxiety spread in the early stage of the novel coronavirus (COVID-19) outbreak in South Korea: structural topic model and network analysis. J Med Internet Res 2020 Jun 02;22(6):e19455 [FREE Full text] [doi: 10.2196/19455] [Medline: 32463367]
7. Bastani P, Bahrami MA. COVID-19 related misinformation on social media: a qualitative study from Iran. JMIR Preprints Preprint posted online March 27, 2020. [FREE Full text] [doi: 10.2196/18932] [Medline: 32250961]
8. Coronavirus disease (COVID-19) advice for the public: Mythbusters. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [accessed 2020-06-10]
9. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
10. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

11. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: cross-sectional online study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
12. Rovetta A, Bhagavathula AS. Global infodemiology of COVID-19: analysis of Google web searches and Instagram hashtags. J Med Internet Res 2020 Aug 25;22(8):e20673 [FREE Full text] [doi: 10.2196/20673] [Medline: 32748790]
13. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
14. Daley DJ, Kendall DG. Epidemics and rumours. Nature 1964 Dec 12;204:1118. [doi: 10.1038/2041118a0] [Medline: 14243408]
15. Bettencourt LMA, Cintrón-Arias A, Kaiser DI, Castillo-Chávez C. The power of a good idea: quantitative modeling of the spread of ideas from epidemiological models. Physica A Stat Mechanics Applications 2006 May;364:513-536. [doi: 10.1016/j.physa.2005.08.083]
16. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
17. Jombart T, Kamvar ZN, Cai J, Pulliam J, Chisholm S, FitzJohn R, et al. reconhub/incidence: Incidence version 1.7.2. Zenodo. 2020. URL: https://zenodo.org/record/3959844#.X8k4sdhKhPY [accessed 2020-08-05]
18. The R Project for Statistical Computing. URL: https://www.R-project.org/ [accessed 2020-08-05]
19. Individuals using the Internet (% of population). The World Bank Open Data. URL: https://data.worldbank.org/indicator/IT.NET.USER.ZS [accessed 2020-10-16]

Abbreviations

WHO: World Health Organization

Edited by G Eysenbach; submitted 18.09.20; peer-reviewed by M Popovic, R Halkes; comments to author 09.10.20; revised version received 30.10.20; accepted 30.11.20; published 15.12.20.

Please cite as:
Nsoesie EO, Cesare N, Müller M, Ozonoff A
COVID-19 Misinformation Spread in Eight Countries: Exponential Growth Modeling Study
J Med Internet Res 2020;22(12):e24425
URL: http://www.jmir.org/2020/12/e24425/
doi:10.2196/24425
PMID:33264102

Original Paper

COVID-19 Symptom-Related Google Searches and Local COVID-19 Incidence in Spain: Correlational Study

Alberto Jimenez Jimenez[1*], MA; Rosa M Estevez-Reboredo[2*], PhD; Miguel A Santed[3*], PhD; Victoria Ramos[4*], PhD

[1]Instituto de Salud Carlos III, Information and Communication Technologies Unit, Madrid, Spain
[2]Instituto de Salud Carlos III, National Epidemiology Centre, Madrid, Spain
[3]Faculty of Psychology, Universidad Nacional de Educación a Distancia, Madrid, Spain
[4]Instituto de Salud Carlos III, Telemedicine and Digital Health Research Unit, Madrid, Spain
[*]all authors contributed equally

Corresponding Author:
Miguel A Santed, PhD
Faculty of Psychology
Universidad Nacional de Educación a Distancia
Calle de Juan del Rosal 10
Madrid, 28040
Spain
Phone: 34 646517577
Email: msanted@psi.uned.es

Abstract

Background: COVID-19 is one of the biggest pandemics in human history, along with other disease pandemics, such as the H1N1 influenza A, bubonic plague, and smallpox pandemics. This study is a small contribution that tries to find contrasted formulas to alleviate global suffering and guarantee a more manageable future.

Objective: In this study, a statistical approach was proposed to study the correlation between the incidence of COVID-19 in Spain and search data provided by Google Trends.

Methods: We assessed the linear correlation between Google Trends search data and the data provided by the National Center of Epidemiology in Spain—which is dependent on the Instituto de Salud Carlos III—regarding the number of COVID-19 cases reported with a certain time lag. These data enabled the identification of anticipatory patterns.

Results: In response to the ongoing outbreak, our results demonstrate that by using our correlation test, the evolution of the COVID-19 pandemic can be predicted in Spain up to 11 days in advance.

Conclusions: During the epidemic, Google Trends offers the possibility to preempt health care decisions in real time by tracking people's concerns through their search patterns. This can be of great help given the critical, if not dramatic need for complementary monitoring approaches that work on a population level and inform public health decisions in real time. This study of Google search patterns, which was motivated by the fears of individuals in the face of a pandemic, can be useful in anticipating the development of the pandemic.

(J Med Internet Res 2020;22(12):e23518) doi:10.2196/23518

KEYWORDS

behavioral epidemiology; big data; smart data; tracking; nowcasting; forecast; predict; infosurveillance; infodemiology; COVID-19

Introduction

Background

During the 2020 Chinese Lunar New Year, massive measures for reducing the spread of the new COVID-19 disease were first enacted by authorities in China [1]. The first reported case of a SARS-CoV-2 infection appeared in late 2019 [2]. Subsequently, the World Health Organization declared the COVID-19 outbreak as a pandemic on March 11, 2020 [3]. The epidemiological characteristics of COVID-19, such as its high transmissibility capacity, virulence, and presence in asymptomatic carriers or those showing only mild symptoms, have yet to be fully understood. At 7 months after the epidemic outbreak in March 2020, we are close to reaching 1 million deaths worldwide [4].

Spain has the fifth highest number of detected COVID-19 cases in the world, behind the United States, Brazil, Russia, and the

United Kingdom [4]. As a consequence, developing a forecasting tool to predict the spread of the epidemic has become critical. This information can help us understand the evolution of COVID-19 and how it affects our health. Furthermore, such information can even be useful in preparing for possible future COVID-19 waves and other pandemics.

Outbreak Detection

In a study by Chu and Qureshi [5], they state the following:

> *Predicting the potential spread of a pandemic like COVID-19 is difficult because we do not have many epidemiological data, such as the transmission mechanism, the contagiousness of the virus, or its mutation patterns, as well as other complex human factors, such as the level of compliance with social distancing measures. Many models recently developed by infectious disease scientists [e.g., the Imperial College model [6]…and The Reich Lab [7]…] can produce vastly different predictions as they are constructed based on various assumptions that may not be close to reality (such as the actual level of compliance with social distancing may be much higher than what is assumed in the model, or the infection rates can vary across different regions and groups of people, which cannot be easily captured by any model).*

Google Trends offers a new approach for potentially predicting changes in the pandemic by tracking individuals' concerns through their searches. Gunther Eysenbach [8] is a pioneer in terms of conducting studies on the use of Google Trends in health approaches. Furthermore, Ginsberg et al [9] found a high correlation between the pattern of web search queries and the percentage of patients with influenza-like symptoms, thereby confirming that at specific points in time, Google Trends can detect influenza expansion 1 or 2 weeks earlier than the Centers for Disease Control and Prevention.

Within the discipline of behavioral epidemiology, there are articles that study fear in the development of epidemics (eg, Epstein et al study [10]). In our study, behavioral factors were summarized through Google searches and then used as a correlation variable to identify patterns in the evolution of epidemics.

In recent years, the number of different search engines that deal with infodemiology—“which studies the determinants and distribution of health information for public health purposes” [8]—are increasing, and Google Trends is being tested as a useful tool for tracking social trends [11]. During the COVID-19 pandemic, we found 7 articles published from up to May 2020 that raised the possibility of predicting the development of the disease (Table 1).

Table 1. Articles that raised the possibility of predicting the development of COVID-19.

Study (Author)	Search engine	Territory	Terms of interest	Time lag
Li et al [12]	Google Trends, Baidu Index, and Sina Weibo Index	China	“Coronavirus” and “Pneumonia”	6-8 days
Husnayain et al [13]	Google Trends	Taiwan	“Handwashing” and “Mask Related Information”	1-3 days
Effenberger et al [14]	Google Trends	China, Republic of Korea, Japan, Iran, Italy, Austria, Germany, United Kingdom, United States, Egypt, Australia, and Brazil	“Coronavirus (Virus)”	11.5 days
Pekoz et al [15]	Google Trends	United States	“Sore Throat,” “Fever,” and Cough	1-2 weeks
Yuan et al [16]	Google Trends	United States	“COVID Pneumonia” and “COVID Heart”	Approximately 12 days
Higgins et al [17]	Google Trends and Baidu Index	China, worldwide data, Italy and Spain, and the US states of New York and Washington	“Shortness of Breath,” “Anosmia,” “Dysgeusia,” “Ageusia,” “Headache,” “Chest Pain,” “Sneezing,” “Diarrhea,” “Fever,” “Cough,” “Nasal Obstruction,” and “Rhinorrhea”	12 days
Lu and Reis [18]	Google Trends	32 countries	“Coronavirus Symptoms,” “Coronavirus Test,” “Fever,” “Cough,” “Coronavirus,” “Runny Nose,” “Dry Cough,” “Sore Throat,” “Chills,” and “Shortness of Breath”	18-22 days

Objectives

In this study, a statistical approach was proposed in order to assess the correlation between the incidence of COVID-19 in Spain and search data provided by Google Trends. We aimed to determine whether Google Trends data that are collected for searches using many different keywords that the public has entered into Google's internet search engine during the COVID-19 outbreak period can predict the number of cases reported by the National Center of Epidemiology in Spain (Centro Nacional de Epidemiología [CNE]).

Methods

Study Design

Our null hypothesis, H0, was as follows: there is no statistically significant relationship between the variables. The proposed alternative hypothesis, H1, was as follows: the obtained correlation coefficient comes from a population whose correlation coefficient is significant.

To achieve the aforementioned objective, we analyzed search data obtained from Google Trends and the official data for the number of daily cases registered by the CNE during the period between February 20, 2020 and May 20, 2020. The rationale behind choosing this time frame for data analysis was that the CNE kept track of the daily cases using a different method until May 20, 2020. The current counting system was implemented on May 11, 2020. The CNE is the government body responsible for collecting and standardizing the data of the 17 autonomous communities that make up the Spanish state. The further analysis of this second dataset has been planned to extend this study.

Variables for Google Trends Search Terms

We were unable to obtain data for sex and gender from the 2 groups of variables used in this study. These 2 datasets were assessed for any correlations. Our methodology did not include data from explicit participants. Although this may have limited our study results, it allowed for the timely interpretation of data.

Google Trends searches were carried out by looking up terms related to COVID-19 symptoms and synonyms for the term "COVID-19." Google Trends provides an index of time series data for the volume of queries users have entered into Google within a given geographic area. Google calculates the number of searches for a given term as a proportion of the total number of searches in each location at any given time. These calculations are normalized with a Google Trends relative search volume (RSV) index, which ranges from 0 to 100. An RSV of 100 designates the date with the highest amount of search activity for a given term [19]. In our previous study [20], we established the mathematical formulation of how Google Trends calculates its monthly RSV index for a particular search term.

The following terms were searched on Google Trends: "cansancio," which translates as "fatigue"; "coronavirus," "COVID 19," "covid 19," and "COVID19"; "diarrea," which translates as "diarrhea"; "dolor de garganta," which translates as "sore throat"; "fiebre," which translates as "fever"; "neumonia," which translates as "pneumonia" and was searched without an accent due to being more relevant; "perdida de olfato," which translates as "lost sense of smell" and was also searched without an accent; and "tos," which translates as "cough."

Possible spurious terms that Google Trends pointed out in its related queries were eliminated in the search strings by putting the negation operator "-" before the spurious term. The full terms searched, along with links to the original Google Trends search pages and negative queries, were as follows: "cansancio -sociedad" [21]; "coronavirus," "COVID 19," "covid 19," and "covid19" [22]; "diarrea" [23]; "dolor de garganta" [24]; "fiebre" [25]; "neumonia" [26]; "perdita olfato" [27]; and "tos -opensigma -rap" [28]. The full original Google Trends search results for terms related to symptoms of COVID-19 were specific to Spain and the period from February 20 to May 20, 2020.

Information Provided by the CNE

The CNE is the official Spanish center that collects and centralizes all the epidemiological information in the country. The accuracy of these data mainly depends on the agencies that supply it. In this study, the data mainly come from the autonomous communities that occupy the second administrative level within the Spanish public administration system.

The CNE itself has given warnings about a certain lack of homogenization in the data coming from the sources of origin (ie, the autonomous communities) during the initial period of the pandemic. This has given rise to certain inconsistencies in the data received, which is a problem that we were able to verify ourselves when transforming the aggregate data from the CNE into daily values. Although we obtained 1 negative value, we did not normalize these data, since we believed that they are indeed representative of the state of emergency we have experienced. In addition, even with this evidently erroneous data, the correlations obtained were very good.

The original initial data for the COVID-19 outbreak in Spain were found on the CNE webpage on May 24, 2020, under the tab "Documentación y Datos" [29]. The official COVID-19 data that the CNE offered at that time were aggregate numbers for the following: polymerase chain reaction (PCR)-positive results, the number of people hospitalized, the number of admissions into intensive care units, and the number of deceased persons. This dataset is no longer available, and the dataset in its place presents the results of the second data collection method, which was implemented on May 11, 2020. Our intention is to do research on this dataset in a future study that uses the same methodology as this study.

In the first part of our analysis, we used PCR-confirmed COVID-19 case data, and for the second part of our analysis, we used 4 separate datasets in order to create daily delay graphs for each symptom or searched term.

Bias, Study Size, and Participants

A potential source of bias may be the choice of Google Trends search terms, as terms for COVID-19 symptoms differ slightly in the Spanish vocabulary. The study sample is representative of the population of Spain, as we restricted the Google Trends data to searches conducted in Spain and the CNE monitors all COVID-19 cases that occur in Spain.

Statistical Analysis

The Pearson correlation coefficient was used to study the linear relationship between 2 continuous variables (ie, each of the symptoms searched in Google Trends vs the number of daily PCR-positive cases). This is a parametric test that infers its results to be representative of the real population, which makes it necessary for the distribution of the sample to resemble that of the real world to ensure the normality of the data. Therefore, the data, which was drawn randomly from a population whose correlated variables are normally distributed, must be validated.

Given that the sample size was <50, the appropriate test for contrasting the goodness of fit to a normal distribution is the Shapiro-Wilk test, in which the null hypothesis implies that the data are normally distributed.

The Pearson correlation coefficient enabled us to understand the intensity and direction of the relationship between variables. It is a symmetric measure; the correlation between x_i and y_i is the same as the correlation between y_i and x_i.

Time lag correlations were measured to assess whether increases in Google Trends data had a correlation with the evolution of the pandemic. A threshold of $P<.05$ was used to determine statistical significance.

Results

Pearson Correlation Analysis and Shapiro-Wilk Test

Tables 2 and 3 present the Pearson correlation coefficients for each of the symptoms categorized in the searches, from the day of the initial search until 21 days later. For each symptom, the day with the highest correlation was noted.

Table 4 shows whether the variables followed a normal distribution.

Table 2. Pearson correlation coefficients and *P* values for each of the symptom-related search terms, excluding fatigue. Time lags with respect to COVID-19 incidence data are also presented[a].

Lag, days[b]	Sore throat		Coronavirus[c]		Fever		Cough		Diarrhea		Pneumonia		Lost sense of smell	
	r	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value
1	0.4848	<.001	0.5090	<.001	0.4009	<.001	0.4149	<.001	0.5870	<.001	0.6551	<.001	0.6770	<.001
2	0.5549	<.001	0.5730	<.001	0.4770	<.001	0.4964	<.001	0.6446	<.001	0.7163	<.001	0.7151	<.001
3	0.6121	<.001	0.6273	<.001	0.5437	<.001	0.5741	<.001	0.6815	<.001	0.7654	<.001	0.7489	<.001
4	0.6708	<.001	0.6273	<.001	0.5437	<.001	0.5741	<.001	0.6815	<.001	0.7654	<.001	0.7489	<.001
5	0.7369	<.001	0.7256	<.001	0.6771	<.001	0.6991	<.001	0.7743	<.001	0.8544	<.001	0.7847[d]	<.001
6	0.7679	<.001	0.7521	<.001	0.7238	<.001	0.7439	<.001	0.7782	<.001	0.8639	<.001	0.7806	<.001
7	0.8055	<.001	0.7740	<.001	0.7656	<.001	0.7794	<.001	0.7780	<.001	0.8756[d]	<.001	0.7726	<.001
8	0.8358	<.001	0.8118	<.001	0.8100	<.001	0.8201	<.001	0.7922	<.001	0.8593	<.001	0.7319	<.001
9	0.8608	<.001	0.8507	<.001	0.8434	<.001	0.8584	<.001	0.7907	<.001	0.8604	<.001	0.6814	<.001
10	0.8743	<.001	0.8766	<.001	0.8751	<.001	0.8822	<.001	0.8031	<.001	0.8501	<.001	0.6356	<.001
11	0.8799	<.001	0.8999[d]	<.001	0.9086[d]	<.001	0.9015[d]	<.001	0.8117[d]	<.001	0.8585	<.001	0.6111	<.001
12	0.8924[d]	<.001	0.8944	<.001	0.9039	<.001	0.8965	<.001	0.7858	<.001	0.8484	<.001	0.5592	<.001
13	0.8672	<.001	0.8468	<.001	0.8788	<.001	0.8681	<.001	0.7001	<.001	0.8127	<.001	0.4968	<.001
14	0.8279	<.001	0.8065	<.001	0.8319	<.001	0.8296	<.001	0.6326	<.001	0.7668	<.001	0.4419	<.001
15	0.7664	<.001	0.7443	<.001	0.7743	<.001	0.7839	<.001	0.5913	<.001	0.7099	<.001	0.3803	<.001
16	0.7214	<.001	0.6811	<.001	0.7234	<.001	0.7448	<.001	0.5192	<.001	0.6415	<.001	0.3259	.004
17	0.6720	<.001	0.6214	<.001	0.6827	<.001	0.7030	<.001	0.4733	<.001	0.5844	<.001	0.2524	.03
18	0.6093	<.001	0.5517	<.001	0.6330	<.001	0.6654	<.001	0.4271	<.001	0.5467	<.001	0.2053	.08
19	0.5788	<.001	0.4838	<.001	0.5810	<.001	0.6161	<.001	0.3607	<.001	0.5142	<.001	0.1690	.16
20	0.5192	<.001	0.4083	<.001	0.5164	<.001	0.5598	<.001	0.2845	.008	0.4694	<.001	0.1140	.34
21	0.4314	<.001	0.3208	.003	0.4349	<.001	0.4921	<.001	0.1975	.051	0.3812	<.001	0.0630	.60

[a] Pearson correlation coefficients and *P* values for each of the symptoms are based on the comparison between Google Trends searches for the term and daily polymerase chain reaction-positive cases.

[b] This column refers to the days of lag between the 2 variables being compared.

[c] "Coronavirus" refers to searches for the following terms: "coronavirus," "COVID 19," "covid 19," and "COVID19".

[d] These are the highest correlations for each symptom.

Table 3. Pearson correlation coefficients and *P* values for the search term “fatigue.” Time lags with respect to COVID-19 incidence data are also presented[a].

Lag, days[b]	Fatigue[c]	
	r	*P* value
22	0.3926	<.001
23	0.3782	<.001
24	0.3632	.001
25	0.4947	<.001
26	0.5296	<.001
27	0.5171	<.001
28	0.5480	<.001
29	0.5253	<.001
30	0.4720	<.001
31	0.5342	<.001
32	0.5016	<.001
33	0.5427	<.001
34	0.5521	<.001
35	0.5664	<.001
36	0.6350[d]	<.001
37	0.4981	<.001
38	0.4711	<.001
39	0.4388	<.001
40	0.4631	<.001
41	0.4915	<.001
42	0.5325	<.001

[a] Pearson correlation coefficients and *P* values for fatigue are based on the comparison between Google Trends searches for the term and daily polymerase chain reaction-positive cases.

[b] This column refers to the days of lag between the two variables being compared.

[c] The searches for “fatigue” correlated less strongly than the searches for the symptoms in Table 1, but they do show stronger correlations after 36 days. Therefore, we presented these results in a separate table with a different scale of days.

[d] This is the highest correlation with regard to searches for “fatigue.”

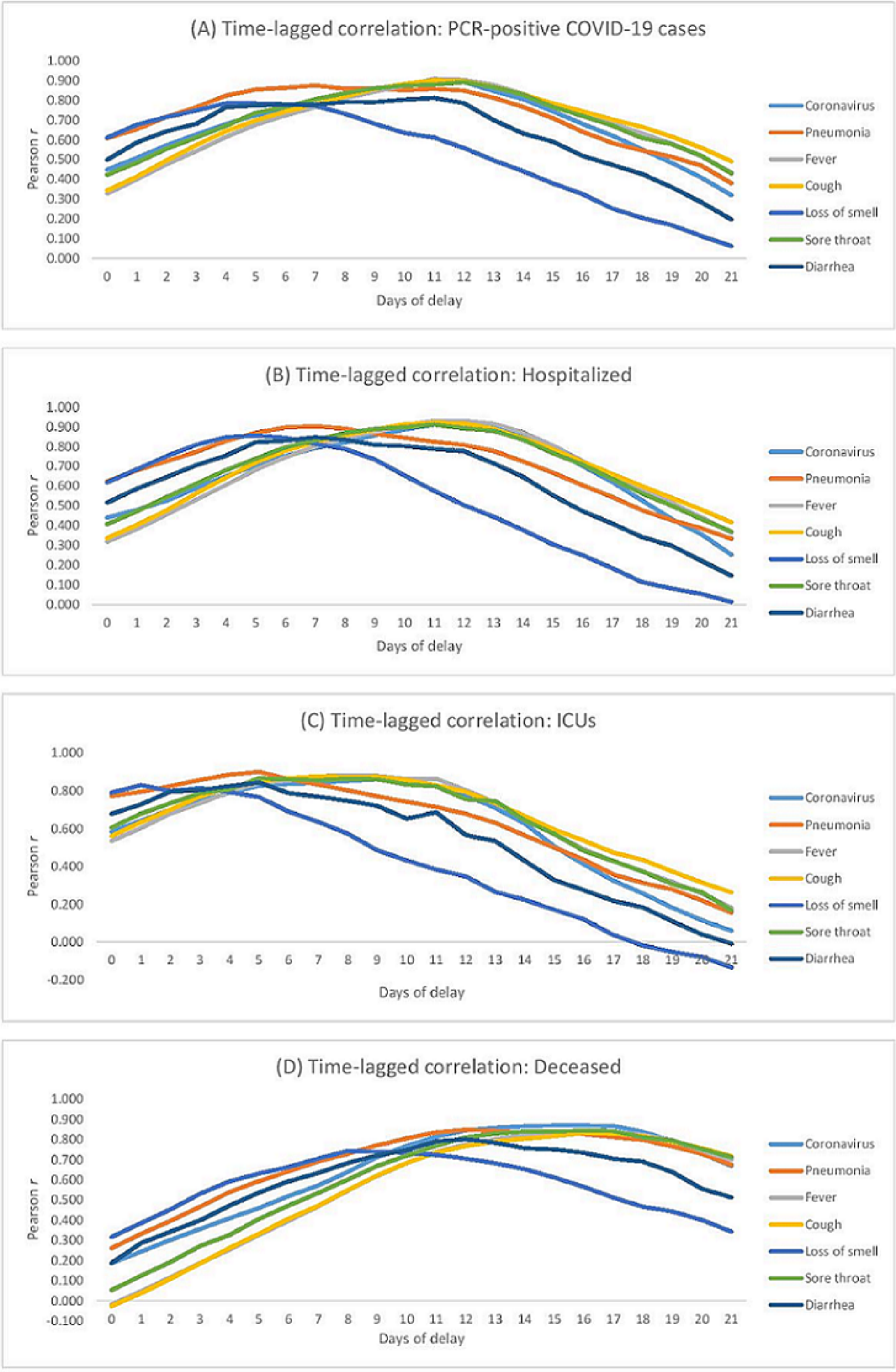
Table 4. Shapiro-Wilk test for normality. The critical region (.908) for a sample size of 21 and significance level of α=.05 was obtained from the critical values of $W_{n,a}$ for the Shapiro-Wilk test.

Variable	Statistic	*P* value
Coronavirus	.945	.28
Pneumonia	.861	.007
Fever	.946	.29
Cough	.944	.26
Lost sense of smell	.885	.02
Sore throat	.929	.13
Diarrhea	.877	.01
Fatigue	.952	.36

According to Table 4, the variables follow a normal distribution. Based on our results regarding searches for the terms “pneumonia”, “loss of smell,” and “diarrhea,” which have a value lower than the critical region (.908) for the 95% confidence level and a *P* value of <.05, the null hypothesis of normality could be rejected. However, these symptoms do follow a linear trend, as can be in Figure 1. Since these values are close to the critical region, they could also be considered to follow a normal distribution, as seen in Figure 2.

Figure 1. Scatter graphs depicting the linear trend in the relationships between daily PCR-positive confirmed COVID-19 cases and GT searches for COVID-19–related keywords in Spain from February to May, 2020. (A) Scatter graph between daily PCR-positive confirmed COVID-19 cases and GT search term Coronavirus. (B) Scatter graph between daily PCR-positive confirmed COVID-19 cases and GT search term Pneumonia. (C) Scatter graph between daily PCR-positive confirmed COVID-19 cases and GT search term Fever. (D) Scatter graph between daily PCR-positive confirmed COVID-19 cases and GT search term Cough. (E) Scatter graph between daily PCR-positive confirmed COVID-19 cases and GT search term Loss of smell. (F) Scatter graph between daily PCR-positive confirmed COVID-19 cases and GT search term Sore throat. (G) Scatter graph between daily PCR-positive confirmed COVID-19 cases and GT search term Diarrhea. (H) Scatter graph between daily PCR-positive confirmed COVID-19 cases and GT search term Fatigue. GT: Google Trends; PCR: polymerase chain reaction.

Figure 2. Graphs that show the day with the best correlation between daily PCR-positive COVID-19 cases and GT searches for COVID-19–related keywords. (A) Graph for the comparison between GT search terms and the number of PCR-positive COVID-19 cases, with different day ranges for delay. (B) Graph for the comparison between GT search terms and the number of people hospitalized due to COVID-19, with different day ranges for delay. (C) Graph for the comparison between GT search terms and the number of admissions to the ICU due to COVID-19, with different day ranges for delay. (D) Graph for the comparison between GT search terms and the number of people deceased due to COVID-19, with different day ranges for delay. GT: Google Trends; ICU: intensive care unit; PCR: polymerase chain reaction.



(A) Time-lagged correlation: PCR-positive COVID-19 cases

(B) Time-lagged correlation: Hospitalized

(C) Time-lagged correlation: ICUs

(D) Time-lagged correlation: Deceased

Outbreak Control Measures

Since the first day of the study period, the number of daily cases began to correlate with searches for all evaluated terms. Since the *P* values for the correlations between daily cases and symptom searches were below the .05 level of significance, it can be stated that the correlation coefficients are significant, which justifies rejecting the null hypothesis. For the majority of the terms that were searched (ie, “coronavirus,” “COVID

19," "covid 19," "COVID19", "fever," "cough," and "diarrhea"), day 11 had the highest correlation with the number of new cases.

With regard to searches for "fatigue," the data began to correlate starting on the third week (ie, day 22). The measures from the previous weeks were eliminated based on the Chauvenet criterion, which states that coefficients outside the confidence interval of 0.3565 and 1.000 can be discarded.

Table 5 shows the coefficients of determination between the number of new cases per day and the rest of the search variables in Google Trends. These coefficients had a critical level of significance (*P*<.001) lower than the established level of significance, which is generally *P*<.05.

Table 5. Coefficients of determination between the number of new polymerase chain reaction-confirmed COVID-19 cases per day and the search variables in Google Trends with a critical level of significance (*P*<.001).

Variable	R^2	*P* value
Coronavirus	0.8098	<.001
Pneumonia	0.7666	<.001
Fever	0.8256	<.001
Cough	0.8128	<.001
Loss of smell	0.6157	<.001
Sore throat	0.7964	<.001
Diarrhea	0.6588	<.001
Fatigue	0.4032	<.001

Table 5 shows that the coefficients of determination were significantly different from 0. Therefore, in this case the null hypothesis is false, as per the Popper methodology [30], and we can affirm that there was a significant positive linear relationship between daily COVID-19 cases and the searches for terms related to COVID-19 and its symptoms on Google Trends. This suggests that the incidence of COVID-19 could be predicted 11 days in advance.

We used graphic procedures to verify the linearity of our results. The graphical representation of the relationship between each term searched and the number of daily cases (Figure 1) shows that there is a linear trend for each relationship. By using the regression lines that were generated, follow-up data can be calculated, including the number of positive cases, thus verifying the correlation between Google Trends searches and the incidence of COVID-19 in Spain.

The graphs in Figure 2 represent the day with the best correlation between both variables (ie, symptom keywords and COVID-19–positive cases). They show the positive relationship between daily cases and the Google Trends searches for terms related to COVID-19 and its symptoms.

Discussion

Study Implications

Much like the influenza virus, SARS-CoV-2 causes ailments with certain flu-like symptoms, such as cough, fever, and fatigue, and in some cases, these symptoms can complicate differential diagnoses. Examining research that uses nontraditional data sources has several implications. We examined the use of search engines for mitigating the impact of the COVID-19 pandemic. Our results demonstrated that Google can potentially be used as a complementary tool to aid in understanding online search behavior, which could help mitigate the adverse effects of the pandemic and expedite the recovery process.

We found that internet search patterns reveal a robust temporal pattern of disease progression for COVID-19. This study shows that internet search patterns can be used to reveal the detailed clinical course of a disease. These data can be used to track and predict the local spread of COVID-19 before widespread laboratory testing becomes available, and help guide current public health responses.

While laboratory testing serves as an important gauge of epidemic spread, it suffers from a number of important limitations. Alternative surveillance approaches are needed to overcome these limitations and serve as a complement to laboratory testing, especially during the critical early stages of a pandemic. Aggregated deidentified internet search patterns have been used to track a wide range of health phenomena and are a potential alternative source of information for surveilling pandemic spread.

When harnessed appropriately, internet search patterns possess a number of powerful advantages over laboratory testing, such as the following: (1) surveillance data are available immediately when a new pandemic emerges, (2) data are available at a population scale in countries with sufficient internet access, (3) delays are minimal, as search data are available the same day, (4) there is no need for individuals to travel to a testing location; people can stay at home, thereby avoiding increased exposure to other people and health care workers, (5) no physical intervention is required, and (6) the data are available for free, independent of the scale of surveillance.

Future research can be focused on checking the progression of symptom-related search terms over time in order to characterize the clinical course of COVID-19 by means of examining a range of possible search term–based definitions for initial symptom onset. This should be based on various combinations of the earliest peaking search terms and a detailed understanding of the stage of illness and the manifestations of COVID-19 in the local environment and over time. Studies have indicated that the spread and severity of COVID-19 can be affected by local conditions, and search volume data can be a valuable complementary tool for studying potential local variations in disease presentation. Given the numerous limitations of laboratory testing, search data are a valuable complementary tool for the population-scale tracking of pandemics in real time.

Principal Results

This study showed that the data obtained from Google Trends searches for Spanish keywords related to COVID-19 (ie, “coronavirus,” “neumonía,” “fiebre,” “tos,” “pérdida de olfato,” “dolor de garganta,” and “diarrea”) correlated with data published by the CNE on the daily incidence of laboratory PCR-confirmed COVID-19 cases, hospitalization, intensive care unit admissions, and deaths from COVID-19, going from R=0.635 for “fatigue” to a maximum of R=0.908 for “fever”. We also found that the Google Trends data correlated with the daily incidence of COVID-19 with an 11-day time lag.

It should be noted that for “fatigue,” the day with the highest correlation was day 36 (ie, the sixth week after the search). Statistically, this is moderately relevant, but considering the high variability in the number of days in incubation, pathogenesis, and generation of an immune response for COVID-19, this relevance may not be so evident when evaluating future COVID-19–positive cases. Consequently, it is possible that fatigue should not be considered as a symptom for assessing and predicting a positive case using Google Trends.

Although we used correlations to examine the possible linear association between search queries and daily COVID-19 incidence, it should be noted that the use of a search engine is voluntary, and self-initiated search queries represent the users who are truly curious or worried about a situation. Thus, we believe that the unobtrusive search behavior of netizens may have resulted in an increase in search volume. The analysis and methods used in this study could aid public health and communication agencies. It is crucial to study this association for the rest of Europe, since other countries, such as Italy, Great Britain, and France, have been affected by the COVID-19 pandemic, and new waves of COVID-19 are foreseeable as long as social distancing measures are relaxed and the winter cold reenters.

This study presents the need for a detailed survey that provides data on the clinical features of COVID-19, prevention strategies, and technological solutions, including search engine data that have been at the forefront of health research. Findings from this study validate and extend previously published works that used Google keywords [5,6,8], and we demonstrate the potential of using Google for monitoring and predicting the evolution of the COVID-19 pandemic. By using Google Trends, this study identified that there is growing interest in COVID-19 worldwide and in countries with a high incidence of SARS-CoV-2 virus infection.

Limitations

Our study used Google Trends, which only provides the search behavior of people using the Google search engine. Future studies should consider studying the same topic, but with other search engine platforms to capture a more diverse population of users. The use of an automated program [31] can improve the accuracy of the data collected and analyzed in countries with a high incidence of SARS-CoV-2 virus infection. Furthermore, the selection of keywords plays a very important role in ensuring the validity of our results. Taking into account that this field of research is relatively new, there is no standard method for reporting, resulting in the same meaning of different terms, different meanings of the same term, and different abbreviations. In addition, search data may be subject to socioeconomic, geographic, or other biases inherent in the local digital divide. Lastly, Google Trends does not provide information about the methods used to generate search data and its algorithms. Therefore, other search engines should be investigated. The transfer of conclusions to countries with a low level of internet access should be done with caution.

Comparison With Prior Work

By using Google Trends, this study identified that there is growing interest in COVID-19 worldwide, as well as in countries with a high incidence of SARS-CoV-2 virus infection. This study is consistent with previous studies, such as those listed in Table 1, since all of those studies found a positive correlation between searches related to COVID-19 and the evolution of the pandemic. Furthermore, the correlation lag model of these studies’ data series is within the range of our findings.

Conclusions

Further research is necessary to determine if the lag detected in our study is related to the results of clinical studies that postulate 97.5% of symptomatic COVID-19 cases develop within 11.5 days after exposure [32]. This 11.5-day adjustment is an improvement over the initial case adjustment date of 15 days. In fact, during the second wave of the pandemic, a 10-day quarantine has been considered sufficient in many places. For visualization purposes, a 10-day moving average would provide slightly clearer plots.

Another priority in the early stages of an emergent pandemic, such as the COVID-19 pandemic, is to characterize the clinical course of symptoms in affected individuals. If population-scale clinical patterns could be ascertained early, it would be beneficial to pandemic tracking, case diagnosis, and treatment. Therefore, we investigated whether internet search data could be used to characterize the clinical course of COVID-19 symptoms over time and provided a search data–based view of the clinical course of the illness.

With regard to future studies, it could be useful to use Pytrends [31], a simple interface for automatically downloading reports from Google Trends. Furthermore, methodologies beyond those first developed [11] must be advanced to approach studies on internet search patterns in a systematic way. For countries where the inflection curve has not yet occurred, a systematic approach can be the most useful if governments monitor the evolution of

Google queries in their country to foresee the best use of their hospital systems.

Acknowledgments

The authors would like to acknowledge the contributions by César Martín Pérez in this study. The authors also highly value the collaboration and support of Amparo Larrauri, Marina Pollán, Pedro Arias, Rosa Cano, and other colleagues from the CNE, who with their hard work in adverse conditions, have provided data on the prevalence of the epidemic in Spain. We would especially like to thank all the workers of the National Health System of Spain who have saved so many lives and who have facilitated control over this pandemic. This work was supported in part by the project PI19CIII/00056 – TMPY 508/19, funding from Sub-Directorate-General for Research Assessment and Health Promotion in Spain (Instituto de Salud Carlos III). The statements made in this study are solely the responsibility of the authors.

Conflicts of Interest

None declared.

References

1. Chen S, Yang J, Yang W, Wang C, Bärnighausen T. COVID-19 control in China during mass population movements at New Year. Lancet 2020 Mar 07;395(10226):764-766 [FREE Full text] [doi: 10.1016/S0140-6736(20)30421-9] [Medline: 32105609]
2. Coronavirus disease 2019 (COVID-19) Situation Report – 94. World Health Organization. 2020 Apr 23. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200423-sitrep-94-covid-19.pdf [accessed 2020-09-06]
3. Coronavirus disease (COVID-19) pandemic. World Health Organization Europe. URL: https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19 [accessed 2020-09-06]
4. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2020-06-10]
5. Chu B, Qureshi S. Predicting the COVID-19 Pandemic in Canada and the US. Ideas. 2020 May 05. URL: https://ideas.repec.org/p/car/carecp/20-05.html [accessed 2020-07-30]
6. Imai N, Dorigatti I, Cori A, Donnelly C, Riley S, Ferguson NM. Report 2: Estimating the potential total number of novel Coronavirus (2019-nCoV) cases in Wuhan City, China. Imperial College London COVID-19 Response Team. 2020 Jan 22. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-update-epidemic-size-22-01-2020.pdf [accessed 2020-09-12]
7. reichlab/covid19-forecast-hub: Projections of COVID-19, in standardized format. GitHub. URL: https://github.com/reichlab/covid19-forecast-hub [accessed 2020-09-12]
8. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006;2006:244-248 [FREE Full text] [Medline: 17238340]
9. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
10. Epstein JM, Parker J, Cummings D, Hammond RA. Coupled contagion dynamics of fear and disease: mathematical and computational explorations. PLoS One 2008;3(12):e3955. [doi: 10.1371/journal.pone.0003955] [Medline: 19079607]
11. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
12. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):2000199 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
13. Husnayain A, Fuad A, Su EC. Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. Int J Infect Dis 2020 Jun;95:221-223 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.021] [Medline: 32173572]
14. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends Analysis. Int J Infect Dis 2020 Jun;95:192-197 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.033] [Medline: 32305520]
15. Pekoz EA, Smith A, Tucker A, Zheng Z. COVID-19 Symptom Web Search Surges Precede Local Hospitalization Surges. SSRN. Preprint posted online on April 30, 2020 . [doi: 10.2139/ssrn.3585532]
16. Yuan X, Xu J, Hussain S, Wang H, Gao N, Zhang L. Trends and Prediction in Daily New Cases and Deaths of COVID-19 in the United States: An Internet Search-Interest Based Model. Explor Res Hypothesis Med 2020 Apr 18;5(2):1-6 [FREE Full text] [doi: 10.14218/ERHM.2020.00023] [Medline: 32348380]

17. Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY, Snot Force Alliance. Correlations of Online Search Engine Trends With Coronavirus Disease (COVID-19) Incidence: Infodemiology Study. JMIR Public Health Surveill 2020 May 21;6(2):e19702 [FREE Full text] [doi: 10.2196/19702] [Medline: 32401211]
18. Lu T, Reis BY. Internet search patterns reveal clinical course of COVID-19 disease progression and pandemic spread across 32 countries. medRxiv. Preprint posted online on September 16, 2020 . [doi: 10.1101/2020.05.01.20087858]
19. FAQ about Google Trends data. Trends Help. URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2020-09-06]
20. Jimenez A, Santed-Germán MA, Ramos V. Google Searches and Suicide Rates in Spain, 2004-2013: Correlation Study. JMIR Public Health Surveill 2020 Apr 13;6(2):e10919 [FREE Full text] [doi: 10.2196/10919] [Medline: 32281540]
21. cansancio -sociedad - Explore. Google Trends. URL: https://trends.google.com/trends/explore?date=2020-02-20%202020-05-20&geo=ES&q=cansancio%20-sociedad [accessed 2020-11-25]
22. coronavirus, COVID 19, covid 19, covid19 - Explore. Google Trends. URL: https://trends.google.com/trends/explore?date=2020-02-20%202020-05-20&geo=ES&q=coronavirus,COVID%2019,covid%2019,covid19 [accessed 2020-11-25]
23. diarrea - Explore. Google Trends. URL: https://trends.google.com/trends/explore?date=2020-02-20%202020-05-20&geo=ES&q=diarrea [accessed 2020-11-25]
24. dolor de garganta - Explore. Google Trends. URL: https://trends.google.com/trends/explore?date=2020-02-20%202020-05-20&geo=ES&q=dolor%20de%20garganta [accessed 2020-11-25]
25. fiebre - Explore. Google Trends. URL: https://trends.google.com/trends/explore?date=2020-02-20%202020-05-20&geo=ES&q=fiebre [accessed 2020-11-25]
26. neumonia - Explore. Google Trends. URL: https://trends.google.com/trends/explore?date=2020-02-20%202020-05-20&geo=ES&q=neumonia [accessed 2020-11-25]
27. perdida olfato - Explore. Google Trends. URL: https://trends.google.com/trends/explore?date=2020-02-20%202020-05-20&geo=ES&q=perdida%20olfato [accessed 2020-11-25]
28. tos -opensigma -rap - Explore. Google Trends. URL: https://trends.google.com/trends/explore?date=2020-02-20%202020-05-20&geo=ES&q=tos%20-opensigma%20-rap [accessed 2020-11-25]
29. COVID-19. Instituto de Salud Carlos III Centro Nacional de Epidemiología. URL: https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos [accessed 2020-05-24]
30. Wilkinson M. Testing the null hypothesis: the forgotten legacy of Karl Popper? J Sports Sci 2013;31(9):919-920. [doi: 10.1080/02640414.2012.753636] [Medline: 23249368]
31. GeneralMills/pytrends: Pseudo API for Google Trends. GitHub. URL: https://github.com/GeneralMills/pytrends [accessed 2020-09-12]
32. Información científica-técnica, enfermedad por coronavirus, COVID-19. Ministerio de Sanidad Centro de Coordinación de Alertasy Emergencias Sanitarias. URL: https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/documentos/ITCoronavirus.pdf [accessed 2020-06-10]

Abbreviations

CNE: Centro Nacional de Epidemiología
PCR: polymerase chain reaction
RSV: relative search volume

Edited by G Eysenbach, R Kukafka; submitted 14.08.20; peer-reviewed by A Husnayain, A Rovetta; comments to author 23.08.20; revised version received 13.09.20; accepted 26.10.20; published 18.12.20.

Please cite as:
Jimenez AJ, Estevez-Reboredo RM, Santed MA, Ramos V
COVID-19 Symptom-Related Google Searches and Local COVID-19 Incidence in Spain: Correlational Study
J Med Internet Res 2020;22(12):e23518
URL: http://www.jmir.org/2020/12/e23518/
doi:10.2196/23518
PMID:33156803

Original Paper

Interrelationships Between Patients' Data Tracking Practices, Data Sharing Practices, and Health Literacy: Onsite Survey Study

Yuhan Luo[1], MS; Chi Young Oh[2], PhD; Beth St Jean[1], PhD; Eun Kyoung Choe[1], PhD

[1]College of Information Studies, University of Maryland, College Park, MD, United States

[2]Chicago State University, Chicago, IL, United States

Corresponding Author:
Eun Kyoung Choe, PhD
College of Information Studies
University of Maryland
4130 Campus Drive
Hornbake Library South
College Park, MD,
United States
Phone: 1 3014051085
Email: choe@umd.edu

Abstract

Background: Although the use of patient-generated data (PGD) in the optimization of patient care shows great promise, little is known about whether patients who track their PGD necessarily share the data with their clinicians. Meanwhile, health literacy—an important construct that captures an individual's ability to manage their health and to engage with their health care providers—has often been neglected in prior studies focused on PGD tracking and sharing. To leverage the full potential of PGD, it is necessary to bridge the gap between patients' data tracking and data sharing practices by first understanding the interrelationships between these practices and the factors contributing to these practices.

Objective: This study aims to systematically examine the interrelationships between PGD tracking practices, data sharing practices, and health literacy among individual patients.

Methods: We surveyed 109 patients at the time they met with a clinician at a university health center, unlike prior research that often examined patients' retrospective experience after some time had passed since their clinic visit. The survey consisted of 39 questions asking patients about their PGD tracking and sharing practices based on their current clinical encounter. The survey also contained questions related to the participants' health literacy. All the participants completed the survey on a tablet device. The onsite survey study enabled us to collect ecologically valid data based on patients' immediate experiences situated within their clinic visit.

Results: We found no evidence that tracking PGD was related to self-reports of *having sufficient information to manage one's health*; however, the number of data types participants tracked positively related to their self-assessed *ability to actively engage with health care providers*. Participants' data tracking practices and their health literacy did not relate to their data sharing practices; however, their ability to engage with health care providers positively related to their willingness to share their data with clinicians in the future. Participants reported several benefits of, and barriers to, sharing their PGD with clinicians.

Conclusions: Although tracking PGD could help patients better engage with health care providers, it may not provide patients with sufficient information to manage their health. The gaps between tracking and sharing PGD with health care providers call for efforts to inform patients of how their data relate to their health and to facilitate efficient clinician-patient communication. To realize the full potential of PGD and to promote individuals' health literacy, empowering patients to effectively track and share their PGD is important—both technologies and health care providers can play important roles.

(J Med Internet Res 2020;22(12):e18937) doi:10.2196/18937

KEYWORDS

consumer health informatics; patient-generated health data; self-tracking; doctor-patient data sharing; health literacy; surveys and questionnaires

Introduction

Background

The prevalence of mobile health apps and wearable devices has enabled patients to track a variety of patient-generated data (PGD) outside the clinic, ranging from biometrics (eg, blood pressure, heart rate) to their everyday activities (eg, exercise, sleep) [1]. Researchers have found that PGD are a valuable source of information for clinicians because they can use the data for diagnosis and treatment [2-6]. Data-driven consultation has the potential to optimize patient care by leveraging effective PGD sharing [7,8]. However, little is known about whether and to what extent patients who track their PGD actually share their data with clinicians in the clinic, which is critical to fully leveraging the potential of PGD to improve health outcomes.

In prior studies that have examined how patients leverage PGD, a key related concept that has often been neglected is health literacy—the cognitive and social skills to gain access to, understand, and use health information in ways that maintain good health [9]. As a multidimensional concept, health literacy characterizes not only an individual's ability to read and understand health information but also the ability to manage one's health and engage with health care providers [9-11]. Although health literacy has been studied in public health for decades, this concept has rarely been examined within the context of studies focusing on PGD tracking and sharing.

Previous studies have found that through self-tracking of their PGD, patients can develop the skills necessary to interpret health information [12] and to communicate with clinicians [13]. In this light, we aim to understand whether patients' self-tracking practices, as a way to help them gain knowledge about their health [14-18], relate to their health literacy and whether patients' health literacy, as an important skill to fully engage with health care providers [10,11,19], relates to patients' practice of sharing their PGD with clinicians.

Related Work

PGD Tracking

According to previous research, PGD is defined as "health-related data created, recorded, or gathered by patients (or by family members or other caregivers) to help address their health concerns" [1,15]. These data can be collected through automated tracking devices (eg, mobile phones, wearable devices), manual journaling, questionnaires, clinic tests (eg, laboratory results), and tests ordered by patients (eg, 23andMe [20]). Researchers have found that tracking PGD is an effective means for personal health management [12,21,22]. For example, people with diabetes who monitor their glucose levels go through several learning phases to develop an understanding of how their daily activities affect their health [12]. Similarly, people with irritable bowel syndrome (IBS), who track their food and symptom data, conduct self-experiments to identify the types of food that trigger their IBS symptoms [22].

Despite its benefits, PGD tracking can be detrimental under some circumstances. For example, tracking calorie intake can exacerbate negative thoughts and behaviors of patients with eating disorders [23]. In this case, clinicians often recommend to such patients that they manage their eating practices by tracking their feelings and reflections about food and by focusing on the positive sides [3,24]. Therefore, when encouraging patients to track and share their PGD, it is important to tailor the tracking plans of individuals to their tracking needs and their specific health conditions.

PGD Sharing

This study was conducted in the clinic where patients were likely to share their PGD with clinicians in person [1]. From the perspective of clinicians, PGD can support in the diagnosis and delivery of personalized treatment [4,6,13]. For example, sleep data are used by clinicians conducting cognitive behavioral therapy for insomnia to inform *sleep prescriptions* [4]. Similarly, food intake data are an important source for diagnosing and treating dietary problems [3]. Although many clinicians acknowledge that having patients share their PGD can improve the quality of care that they can provide [4,6,15,25-27], their views on sharing PGD are not always positive because of their increased liability, lack of time, doubts regarding data accuracy, integration difficulties, and security concerns [6,28-31].

Patients' preferences for sharing PGD also vary: some patients consistently share their data to keep their clinicians informed [1], some resist sharing their data because of privacy concerns or fear of being judged [3], and others share their PGD, but only to the extent to which they trust their clinicians [4,32]. Although researchers have examined patients' data tracking and sharing practices, we lack empirical data on whether and to what extent patients actually share their self-tracking data during clinical visits and the factors influencing their data sharing practices.

Health Literacy

Health literacy has been declared a national priority in the United States [33]. Previous studies have found that patients with higher health literacy are more likely to be informed about their health [34], to engage in healthy behaviors [35,36], and to feel comfortable while communicating with health care professionals [37,38]. Therefore, health literacy has been used as a critical measure for evaluating the effectiveness of health care technologies [18,39].

To assess patient health literacy, researchers have developed various instruments. The most widely used measures include the Test of Functional Health Literacy in Adults (TOFHLA) [40], the eHealth Literacy Scale (eHEALS) [41], and the Health Literacy Questionnaire (HLQ) [11]. The TOFHLA captures the reading comprehension and numeracy of patients within health care contexts [40], whereas the eHEALS measures the skills of patients related to finding, evaluating, and applying electronic health information to health problems [41]. We chose the HLQ because it captures multiple independent constructs regarding different aspects of health literacy that can yield actionable insights [42,43]. According to the grounded psychometric development and validation of the HLQ [11], each of the 9 metrics in HLQ provides unique insights into 9 areas of health literacy and therefore can be used separately when all the questions in 1 metric are asked (each metric has 4-5 questions). To keep the survey to a reasonable length, we included 2 of the

9 constructs included in the HLQ—*having sufficient information to manage one's health (HSI)* and *the ability to actively engage with health care providers (AE)*—that are particularly relevant to PGD tracking and sharing [11]. Individuals with higher levels of HSI feel *confident that they have information they need to live with and manage their health*; those with lower levels of HSI feel they lack *knowledge in managing their health*. Individuals with higher levels of AE are *proactive about their health and feel in control in relationships with health care providers*; those with lower levels of AE are *passive in their approach to health care and do not have a sense of agency in interactions with clinicians* [11].

Research Questions

Although health literacy has not yet been explicitly considered in studies related to PGD, researchers have found that tracking PGD can increase people's awareness of their health [14] and enable them to better answer their clinicians' questions [15,44,45]. Therefore, we speculate that tracking PGD may relate to one's perception regarding HSI and AE. In addition, we assume that patients who shared their PGD in the clinic had tracked their PGD before and were equipped with sufficient HSI and AE. For example, a high level of HSI could enable patients to share their health data with clinicians by equipping patients with sufficient knowledge about their health and a high level of AE could facilitate patients' data sharing as a way of engaging with clinicians. Thus, we investigate the following research questions (RQs):

- RQ1: Do patients' self-tracking practices relate to their HSI and AE?
- RQ2: Do patients' self-tracking practices, along with their HSI and AE, relate to their practices of sharing their PGD with their clinicians?
- RQ3: What benefits and barriers do patients encounter with regard to sharing their PGD with their clinicians?

Although most PGD-related studies have examined patients' experiences retrospectively after some time had passed since their clinic visit [1,13], we sought to gather the immediate perceptions of patients regarding their experiences of meeting with a clinician. We surveyed 109 patients who had just met with a clinician at a university health center (UHC), asking them about their practices of tracking and sharing PGD and assessing their health literacy (ie, HSI and AE). Conducting the survey onsite at the UHC enabled us to collect the experiences of participants from their current clinic visit, which helped reduce recall bias and enabled us to gather ecologically valid data.

Methods

Survey Design and Measures

Throughout the survey, we used the term *personal health data* instead of PGD to aid participants' understanding. At the beginning of the survey, we defined *personal health data* as *health-related data that people keep track of about themselves, such as exercise (eg, step count, miles run), sleep, diet, heart rate, and blood pressure*. Table 1 and Textbox 1 describe our survey measures, including how we defined and assessed each construct.

Table 1. Quantitative survey measures.

Quantitative measures and descriptions	Scale or categories
Health status	
Self-reported health status of participants	1: very poor to 5: excellent
Tracking practices	
Number of data types: the total number of different types of health data participants reported being tracked	0: track nothing to 9: track 9 types of PGD[a]
Tracking group: how participants kept track of their PGD	• Nontrackers: tracked nothing • Memory trackers: tracked PGD relying solely on their memory • Tool trackers: tracked PGD using at least one type of tool (eg, paper, mobile app)
Sharing practices	
Data sharing: whether participants had shared their health data with clinicians during their clinic visit	Yes or no
Willingness to share: how likely participants said they would be willing to share their data with clinicians in the future	1: very unlikely to 5: very likely
Health literacy	
Having sufficient information to manage one's health	Averaged score from 4 measures on a scale of 1: *strongly disagree* to 4: *strongly agree* (Multimedia Appendix 1: questions 16-19) [11]
The ability to actively engage with health care providers	Averaged score from 5 measures on a scale of 1: *cannot do* to 5: *very easy* (Multimedia Appendix 1: questions 20-24) [11]

[a] PGD: patient-generated data.

Textbox 1. Qualitative survey measures.

- **Data tracked:**

  The types of patient-generated data (PGD) that participants reported tracking (multiple-choice question with an option to write in additional data types)
- **Data shared:**

  The types of PGD that participants reported sharing during the particular appointment with the clinician (multiple-choice question with an option to write in additional data types)
- **Data wished to track and be shared:**

  The types of PGD that participants thought would have been helpful if they had tracked and shared them (open-ended question: "List up to 3 types of personal health data that you think it would have been helpful if you had tracked them and shared them with your doctor today")
- **Sharing benefits:**

  Benefits that patients perceived they would have had by sharing their PGD during the particular encounter (an open-ended question asking participants to explain how sharing particular types of PGD would have been helpful)
- **Sharing barriers:**

  Barriers that patients encountered in sharing their PGD during their visit to the clinic (multiple-choice question: "what barriers (if any) did you have when sharing your PGD with your doctor?" and a follow-up question asking for further explanation)

In addition, we collected participants' (1) demographic information (eg, age, gender, occupation, educational attainment, household income, first language), (2) access to technology (eg, internet, mobile phone), and (3) details of their meeting with the clinician (eg, reasons for visit, satisfaction). The survey consisted of 39 questions (Multimedia Appendix 1) and was administered using Qualtrics [46] on a tablet device we provided.

Data Collection

Data collection for this study was part of a larger project that involved onsite surveys and follow-up interviews at a UHC. This paper focuses on the survey results. Following the UHC's guidance, 3 researchers set up a study space next to several clinics, including primary care, women' s health, behavioral health, immunization or allergy, and alternative medicine. We approached patients as they exited clinic offices and asked whether they had just met with a clinician. If the patient answered *yes*, we invited them to participate in the survey. If they expressed an interest, we then led them to the study space, secured their informed consent, and instructed them to take the survey using a tablet paired with a keyboard. As we had 2 tablets set up in the study space, we were able to host 2 participants at the same time. During the study, 2 to 3 researchers were sitting behind the tablets, and each participant was sitting in front of each tablet to complete the survey. Participants would inform the researchers when they completed the survey. We did not recruit any patients who had visited the UHC for mental health reasons, as per the UHC's request.

The survey took participants about 4 to 16 minutes to complete (mean 7.60, SD 2.62). Each participant received a US $5 campus dining gift card after completing the survey. The study was approved by the Institutional Review Board at the researchers' university.

Data Analysis

To answer RQ1 ("Do patients' self-tracking practices relate to their HSI and AE?"), we used multivariate multiple regression to examine whether the number of data types participants tracked and their tracking group (ie, nontracker, memory tracker, or tool tracker) were related to their HSI and AE levels. As patients' health status has been found to highly correlate with their health literacy [34], we controlled their health status as a covariate. In addition, we dummy coded the tracking group using nontrackers as the reference group.

To answer RQ2 ("Do patients' self-tracking practices, along with their HSI and AE, relate to their practices of sharing data with doctors?"), we used multiple logistic regression to predict whether participants had shared their PGD during a particular visit based on their HSI, AE, the number of data types they tracked, and their tracking group. Using this same set of predictors in addition to whether participants had shared their data, we used multiple linear regression to predict participants' willingness to share their data with clinicians in the future.

To answer RQ3 ("What benefits and barriers do patients encounter with regard to sharing PGD with doctors?"), we first analyzed participants' responses to the questions regarding their perceived benefits and barriers of sharing PGD with clinicians, and then categorized these responses using a bottom-up (inductive) approach. In this manner, we developed a coding dictionary that reflected different sharing-related benefits and barriers. We later reorganized the initial codes into potential themes, which complemented our findings for RQ2 by explaining why participants had or had not shared their health data with their clinician during the particular clinic visit.

Results

Participants

In total, 112 patients participated in our study; however, 3 patients were subsequently excluded from our analysis as they had visited the UHC simply to receive a flu shot, fill out a form, or get a massage. Among the remaining 109 patients, 86 (78.9%) were female and 73 (67.0%) were native English speakers, and their ages ranged from 18 to 37 years (mean 21.0, SD 3.82).

Most of the participants were college students (77/109, 70.6%), some (18/109, 16.5%) were full-time employees, and the rest (14/109, 12.8%) were unemployed. Except for 2, all participants (107/109, 98.2%) had access to the internet and a mobile phone. Participants reported visiting the clinic for an inquiry or examination for a specific injury, illness, or condition (65/109, 59.6%), a follow-up visit (22/109, 20.2%), a regular checkup (18/109, 16.5%), or something else (4/109, 3.7%).

Of the total, 87.2% (95/109) of participants reported tracking at least one type of PGD. Of these 95 participants, 38 (35%) were *memory trackers* who tracked their health data relying on their memory only and 57 (52%) were *tool trackers* who tracked their data using at least one tool (eg, paper journal, mobile app, website). The remaining 12.8% (14/109) of participants were *nontrackers*, who did not track any PGD. Table 2 provides descriptive statistics for each of the 3 tracking groups.

Table 2. Descriptive statistics for the 3 tracking groups: nontrackers, memory trackers, and tool trackers (the scale of having sufficient information to manage one's health ranges from 1 (strongly disagree that one has sufficient information to manage their health) to 4 (strongly agree that one has sufficient information to manage their health) and the score of one's ability to actively engage with health care providers ranges from 1 (cannot actively engage with health care providers) to 5 (very easy to actively engage with health care providers), with higher scores indicating greater health literacy; N=109).

Participants' information	Total or overall mean values	Nontrackers (n=14; 12.8%)	Memory trackers (n=38; 34.9%)	Tool trackers (n=57; 52.3%)
Age, years; mean (SD)	21.0 (3.82)	20.0 (2.70)	20.5 (3.92)	21.0 (4.70)
Gender, n (%)				
Female	86 (78.9)	11 (78.6)	27 (71.1)	48 (84.2)
Male	23 (21.1)	3 (21.4)	11 (28.95)	9 (15.8)
Health status, n (%)				
Excellent	13 (11.9)	2 (14.3)	7 (18.4)	4 (7.0)
Very good	27 (24.8)	2 (14.3)	7 (18.4)	18 (31.6)
Good	50 (45.9)	7 (50.0)	17 (44.7)	26 (45.6)
Fair	18 (16.5)	3 (21.4)	6 (15.8)	9 (15.8)
Poor	1 (0.9)	0 (0.0)	1 (2.6)	0 (0.0)
Reason for visit, n (%)				
Inquiry or examination	65 (59.6)	8 (57.1)	25 (65.8)	32 (56.1)
Follow-up visit	22 (20.2)	2 (14.3)	7 (18.4)	13 (22.8)
Regular checkup	18 (16.5)	3 (21.4)	5 (13.2)	10 (17.5)
Other	4 (3.7)	1 (7.1)	1 (2.6)	2 (3.5)
Number of data types tracked, mean (SD)	3.46 (2.3)	0 (0)	3.53 (2.0)	4.26 (2.1)
Was data shared during the appointment? n (%)				
No	16 (14.7)	3 (21.4)	5 (13.2)	8 (14.0)
Yes	93 (85.3)	11 (78.6)	33 (86.8)	49 (85.96)
Number of data types shared, mean (SD)	2.50 (2.01)	2.64 (2.47)	2.21 (1.38)	2.67 (1.95)
Likely to share data with clinician in the future? n (%)				
Very likely	61 (56.0)	11 (78.6)	22 (57.9)	28 (49.1)
Somewhat likely	31 (28.4)	0 (0.0)	13 (34.2)	18 (31.6)
Neutral	11 (10.1)	2 (14.3)	1 (2.6)	8 (14.0)
Somewhat unlikely	3 (2.8)	0 (0.0)	1 (2.6)	2 (3.5)
Very unlikely	3 (2.8)	1 (7.1)	1 (2.6)	1 (1.8)
Health literacy scores, mean (SD)				
HSI[a] (Cronbach α=.86)	3.08 (0.53)	3.11 (0.43)	2.97 (0.57)	2.97 (0.58)
AE[b] (Cronbach α=.88)	4.16 (0.55)	4.20 (0.54)	4.16 (0.53)	4.15 (0.61)

[a] HSI: having sufficient information to manage one's health.

[b] AE: the ability to actively engage with health care providers.

RQ1: Relationship Between Self-Tracking Practices and Health Literacy of Participants (HSI and AE Scores)

The regression model that predicted HSI based on the participants' health status, the number of data types they tracked, and their tracking group was significant. However, there were no significant effects of the number of data types tracked or the participant tracking group (Table 3). Using the same independent variables, the regression model that predicted AE was found to be significant, and there was a significant main effect of the number of data types participants tracked. This result indicates that when controlling for participants' health status and their tracking group, those who tracked more data types tended to have higher AE.

Table 3. The multivariate linear regression models predicting having sufficient information to manage one's health and one's ability to actively engage with health care providers (the tracking group was dummy coded using nontrackers as the reference group). The model that predicts HSI achieved a power of >0.90, and the other model that predicts AE achieved a power of >0.80. Both models have been tested to ensure the absence of multicollinearity (Variance Inflation Factor <3.69).

Outcome variable	HSI[a]	AE[b]
Adjusted R^2	0.146	0.106
Effect size	0.171	0.119
***F* ratio**	**5.618**[c]	**4.205**[d]
P value	<.001	.002
Health rating	**0.225**[c]	**0.200**[c]
P value	<.001	<.001
Number of data types	**−0.004**	**0.056**[e]
P value	.87	.04
Memory trackers	**−0.154**	**−0.279**
P value	.38	.16
Tool trackers	**0.032**	**–0.323**
P value	.86	.11

[a] HSI: having sufficient information to manage one's health.

[b] AE: the ability to actively engage with health care providers.

[c] *P*<.001.

[d] *P*<.01.

[e] *P*<.05.

RQ2: Relationships Between PGD Tracking, Health Literacy (HSI and AE Scores), and PGD Sharing in the Clinic

Tracking Does Not Always Lead to Sharing

A total of 85.3% (93/109) of our participants, including 11 nontrackers, reported that they had shared their PGD with their clinicians during the particular visit. The logistic regression predicting whether participants had shared their data during the particular visit was not significant (Table 4). This result indicates that whether participants shared their PGD during the visit was unrelated to the number of data types they tracked, their tracking group, and their HSI and AE levels.

Table 4. The multiple logistic regression model predicting whether a participant had shared their patient-generated data during the particular clinic visit. The model achieved a power of >0.75 and has been tested to ensure the absence of multicollinearity (Variance Inflation Factor <3.79).

Outcome variable	Data sharing
Nagelkerke R^2	0.105
***F* ratio**	**1.277**
P value	.24
HSI[a]	**0.089**
P value	.88
AE[b]	**−1.175**
P value	.06
Number of data types	**0.264**
P value	.12
Memory trackers	**−0.263**
P value	.78
Tool trackers	**−0.608**
P value	.55

[a] HSI: having sufficient information to manage one’s health.

[b] AE: the ability to actively engage with health care providers.

Health Literacy and Willingness to Share One’s PGD With Clinicians in the Future

The regression model predicting participants’ willingness to share their PGD with clinicians in the future was significant, and there were significant effects of whether participants had shared their PGD during the particular visit and their AE level (Table 5). This indicates that, when controlling for other variables, participants who shared their data during the visit and those with a higher level of AE were more likely to be willing to share their PGD with clinicians in the future.

Table 5. The multiple linear regression model predicting a participant's willingness to share their patient-generated data in the future. The model achieves a power of >0.95 and has been tested to ensure the absence of multicollinearity (Variance Inflation Factor <3.80).

Outcome variable	Sharing willingness
Adjusted R^2	0.178
Effect size	0.217
***F* ratio**	**4.888**[a]
P value	<.001
Data sharing	**0.812**[b]
P value	.001
HSI[c]	**0.277**
P value	.14
AE[d]	**0.482**[b]
P value	.009
Number of data types	**0.046**
P value	.31
Memory trackers	**−0.182**
P value	.57
Tool trackers	**−0.442**
P value	.19

[a] *P*<.001.

[b] *P*<.01.

[c] HSI: having sufficient information to manage one's health.

[d] AE: the ability to actively engage with health care providers.

RQ3: Benefits of, and Barriers to, Sharing One's PGD With Clinicians

When asked to list up to 3 types of PGD that the participants felt would have been helpful if they had tracked and shared with their clinicians, 72.5% (79/109) of our participants provided valid responses. These responses included body measures (eg, blood pressure, glucose level, heart rate, body temperature; 47/79, 60%), food (36/79, 46%), sleep (27/79, 34%), water (25/79, 32%), exercise (19/79, 24%), stress (7/79, 9%), menstrual cycle (6/79, 8%), symptoms (6/79, 8%), mood (3/79, 4%), and others (5/79, 6%). In the subsequent open-ended question, participants explained why these data would have been helpful if they had tracked and shared them: (1) to better inform clinicians in detail about their health condition; (2) to receive more personalized care; and (3) to better articulate their illness and health concerns (Textboxes 2 and 3).

Textbox 2. Benefits of sharing one’s patient-generated data with clinicians.

- Better inform clinicians of their health condition: Participants felt that sharing their patient-generated data (PGD) is generally helpful for clinicians to understand their health condition and identify any abnormality
  - “More info, better diagnosis” (P29)
  - “Sharing my body temperature might have helped doctors identify abnormality.” (P48)
  - “Mood, sleep, and stress could have been helpful because [the doctor] would understand my illness severity.” (P103)
- Enable clinicians to deliver more personalized care: Participants stated that clinicians could provide better advice or treatments that are more relevant to them
  - “If I had tracked and shared my skincare routine, the doctor can offer more personalized advices for me.” (P15)
  - “The doctor could have told me if my current numbers [food, water, sleep] were healthy so [I] could adjust [them] accordingly.” (P23)
- Help patients to better articulate their illness and health concerns: Participants who experienced difficulties communicating with their clinicians believed that sharing concrete health data could help them better articulate their symptoms and concerns with evidence
  - “I could have proved that I’ve been having fevers for the past couple of days, in addition to not eating, drinking, and sleeping as much as usual.” (P46)
  - “Instead of saying ‘I have trouble sleeping until 4am and barely sleep,’ I could have said ‘in the past month I have gotten an average of X hours of sleep per night.’” (P66)

Textbox 3. Barriers to sharing one’s patient-generated data with clinicians.

- Uncertainty regarding the relevance and usefulness of their data: Participants were not sure if the data they had tracked was relevant, important, and useful in relation to their current health issues
  - “I am not sure if some symptoms are relevant to my headache and I don’t want to waste time.” (P82)
  - “Sometimes I don't know what aspects of my health are important to share with my doctor in regards to the health issue I'm having at the moment.” (P102)
- Perceived irrelevance of one’s data: Participants considered that their patient-generated data (PGD) was not relevant to their current health concerns
  - “My problem is women health problem; therefore, the data is not very relevant and she did not ask either.” (P22)
  - “Some data, such as my exercise and sleep schedule, is not relevant.” (P64)
- Not having enough data: Participants were unable to share their PGD because they had not sufficiently tracked their data
  - “My blood pressure and heart rate vary quite a bit. I would love to be able to track this but I don’t have a good device to do so.” (P65)
  - “I am somewhat not precise with what I tell them, so that could mess with the data they need to help me.” (P106)
- Limited time to spend with their clinician: Participants expressed concerns regarding the limited time they got to spend with their clinicians
  - “The time I spend with the doctor is so limited and it doesn’t seem like the doctor found it necessary to ask any more additional information.” (P46)
  - “I feel like doctors never have enough time. They’re always in a rush.” (P65)
- Privacy concerns: Participants were concerned about talking to clinicians they were not familiar with and about the privacy of their health data
  - “I feel uncomfortable sharing personal info with people I just met.” (P95)
  - “Hacking of computer systems makes me very concerned about my privacy.” (P34)
- Fear of being judged: Participants worried about doctors judging them because of unhealthy behaviors
  - “I say less than what I have been doing so that I am not get frowned upon.” (P45)
  - “I am afraid of sharing my smoking habits with doctors.” (P52)

When asked about the barriers they had encountered with regard to sharing their PGD with the clinician, 42.2% (46/109) of our participants reported that they did not experience any barriers, whereas the other 57.8% (63/109) of our participants described specific barriers. Their barriers included uncertainty about the relevance and usefulness of their data (18/63, 29%), feeling their data were irrelevant (14/63, 22%), lack of adequate data (12/63, 19%), having limited time to spend with the clinician (11/63, 18%), privacy concerns (7/63, 11%), fear of being judged (4/63, 6%), and others (5/63, 8%). In Textboxes 2 and 3, we list and provide descriptions and sample quotes for each of these barriers.

Discussion

Reflecting on the Relationships Between PGD Tracking and Health Literacy

Our results show that neither the number of data types that participants tracked nor how they tracked their data relates to their HSI; however, when more types of PGD participants were tracked, their AE tended to be higher. We suspect that as participants tracked more types of PGD, they had more information about their health in mind. Armed with this information, participants could more actively engage with their clinicians, articulate their health concerns, and ask and answer questions regarding their health. However, tracking more types of data did not necessarily transfer to useful *knowledge* that participants felt helped them to manage their health. Many participants mentioned that they were not sure about the relevance and usefulness of their data.

Although relatively rare, some researchers have examined the role of health literacy in the ability of individuals to track and monitor their health behaviors [47,48]. For example, Porter et al [47] conducted a diary study focused on people's recording of their daily physical activities and found that highly health literate individuals were able to achieve a higher accuracy in their diaries. Zoellner et al [48] significantly reduced the sugar-sweetened beverage intake of the participants using a health literacy–centric intervention. Therefore, health literacy is seen as an essential *human tool* to support individuals in taking advantage of self-tracking technologies [49]. Although these studies did not examine how to improve the health literacy of individuals, we see opportunities to improve health literacy in people through their engagement in the very act of self-tracking itself and through feedback generated based on one's personal health data. One opportunity is to deliver health knowledge relevant to individuals' PGD. For example, food tracking apps, such as MyFitnessPal [50], suggest whether one's food intake meets their nutritional needs when they record an entry and provide recommended dietary guidelines with feedback on one's food consumption. Unlike traditional health literacy programs that focus on delivering general health knowledge [39,51], specific information related to an individual's own health data can be easier to understand because it is tied to their own health concerns. However, we must keep in mind that interpretation of automatically tracked data is sometimes unreliable or may even cause misunderstandings [52,53]. For example, sleep tracking apps using pressure sensors to estimate sleep hours can cause false-positive detections when people are performing other activities in bed [53]. Therefore, in delivering health information based on an individual's data to promote health literacy, it is important to ensure the accuracy of the data and to provide relevant and actionable advice.

The Gaps Between Tracking and Sharing PGD

Our results show that patients who track their data do not always share those data with clinicians. Patients who shared their data during their clinic visits and those with a higher level of AE reported greater willingness to share their PGD with clinicians in the future. Although participants acknowledged several benefits of sharing their data with clinicians, over half of the participants reported one or more barriers to data sharing. First, participants were uncertain about the relevance of the data to their health or felt that their data were not relevant. Although not all types of PGD are necessarily related to patients' immediate health concerns, these data can be important health indicators for preventative care [54-56]. For example, an individual's activity level and caffeine consumption can affect their sleep quality [54]; similarly, changes in one's heart rate may indicate a risk of cardiovascular disease [56]. To optimize the value of PGD, it is important to empower patients to share their data with clinicians when necessary. We believe that both people and technology play important roles in this process: clinicians can point out specific types of PGD that may be helpful for diagnosis and decision making; tracking tools can be designed to inform patients of how their PGD may or may not be relevant to their present health concerns. For example, by supporting self-experimentation, tracking tools can draw potential correlations between patients' health indicators (eg, symptoms) and their daily behaviors [22]. These tools can also provide feedback based on the data and other contextual information of individuals to help them reflect on their behaviors and better assess the relevance of their data to their health [16].

Second, participants reported that sharing their PGD with clinicians was sometimes difficult because they did not have enough data. Some participants worried that their data were imprecise, whereas others lacked access to professional devices to measure important body metrics, such as heart rate and blood pressure. To encourage patients to share their PGD with clinicians, we argue that there is a need to better support people's ability to capture sufficient and accurate PGD over time. Due to the high data capture burden [14] and lack of flexibility in existing tracking tools [13,14], many people cannot adhere to the tracking plans, and may even abandon tracking. To address these challenges, we call for flexible and low-burden tracking tools to better support the ability of patients to collect their health data over a long period, which can help both patients and clinicians. For example, semiautomated tracking tools can be used to lower data capturing burden by leveraging automatic tracking (eg, sleep data from Fitbit [57]) and enable flexible data capture through manual tracking (eg, mood) [58,59]. In addition, efforts need to be made to lower the barriers to tracking devices and to make them more accessible to the public (eg, setting up blood pressure kiosks).

Third, participants were reluctant to share their data because of the limited time they had to meet with their clinician. Participants explained that to make the best use of the time, they did not want to waste the time on sharing their PGD, especially when they considered that their data would not help clinicians. However, because patients are not medical experts, they could not always tell which data are relevant to their health, especially if they kept track of various types of data. Although our work did not examine clinicians' perspectives on PGD sharing, previous work has found that clinicians are also concerned about the time they need to review PGD in the clinic [6,60] and may even consider these data to be distracting [52]. To resolve the dilemma of clinicians having limited time and patients feeling uncertain about whether their personal health data can be helpful, an approach should be developed to help patients share their data efficiently. For example, Fitbit recently rolled out a health and wellness report [61] that includes a visual or textual summary of a user's data, which the individual can print out and take to the clinician's office [4]. Although the relevance of these data to one's health may be unclear to the patient, we believe that this is a promising step to support patients in sharing their data with their clinicians, leveraging consumer health technologies.

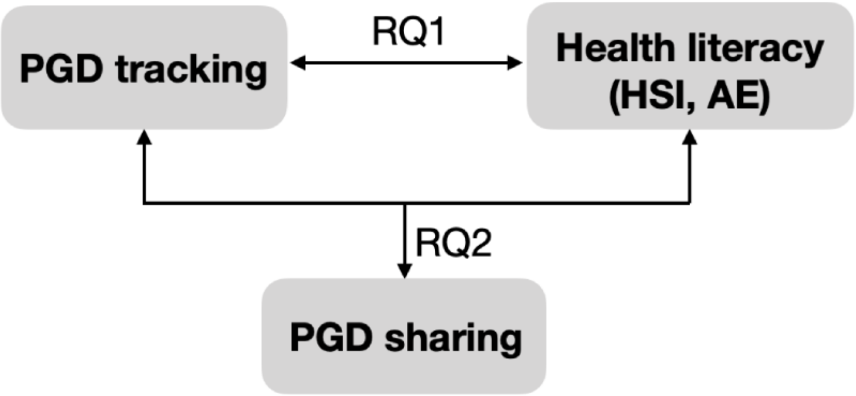
Finally, some participants reported that they were reluctant to share their PGD because of privacy concerns and fear of being judged, especially when asked to share data relating to unhealthy behaviors, such as smoking. This fear may stem from how tracking tools often pass judgment on people's data, as researchers have noticed many self-tracking tools (eg, MyFitnessPal [50], Lose It! [62]) posit weight gain and extra calorie intake as negative outcomes (eg, using red to highlight these data points). Although such designs serve as an alert for individuals to maintain a healthy lifestyle, they can increase the anxiety of those who have weight concerns or dietary problems [63]. To help patients overcome their fear of being judged, tracking tools should avoid creating an environment that perpetuates stigma and that passes explicit or implicit judgments on users.

Limitations and Future Work

This study has limitations that should be considered when interpreting the results and implications. First, although significant, the adjusted R-square statistics in our regression models were relatively low, with a medium effect size (Table 3), indicating that the models explained only a small proportion of the variances in our outcome variables (ie, HSI, AE, willingness to share one's data with clinicians in the future). This could be partly because of either (1) omitting other confounding factors that could influence the health literacy of participants (eg, socioeconomic status, health insurance) [64,65] and data sharing practices (eg, the questions clinicians asked) or (2) sampling bias resulting from recruiting mostly young, healthy, and tech-savvy college students with adequate technology access and higher HSI and AE than patients in general [66,67]. Second, we characterized the tracking practices of participants based on the number of data types they tracked and how they tracked their data, which did not capture all dimensions related to self-tracking. Third, some participants reported sharing PGD they had not tracked (eg, blood pressure taken in the clinic), possibly because (1) they interpreted PGD as encompassing health-related data captured in the clinic, despite our best efforts to define and communicate its meaning or (2) they were able to retrospectively recall or estimate some forms of PGD, even if they had not explicitly tracked the data (eg, weight). Fourth, to make the survey easy and quick to respond to, we omitted questions inquiring about how the PGD sharing happened during the clinic visit. For example, we did not specifically ask participants what triggered their data sharing (eg, whether participants were asked by the clinician, had proactively shared their data themselves, or had their body metrics assessed in the clinic). In addition, we did not ask participants how they shared their data (eg, verbal communication, screen sharing). Finally, we conducted the survey at a UHC, where the clinicians might not be the primary physicians of the participants, which may limit the generalizability of our findings.

Despite these limitations, we believe that, in optimizing health care for the next generation, investigating the interrelationships between PGD tracking and sharing practices, health literacy, and clinic experiences of tech-savvy college students is an important starting point. Going forward, we aim to further examine whether and how health literacy relates to different PGD tracking and sharing contexts—for example, replicating the study with different patient populations, such as those with more severe health conditions and those with low health literacy. Other dimensions of patients' self-tracking practices, such as the frequency with which they track their data (eg, routine vs casual trackers), and triggers and methods of data sharing in the clinic warrant future research [68]. In addition, our study focused only on 2 health literacy constructs (HSI and AE). It is worth investigating how patients' data tracking and sharing practices may relate to other health literacy constructs, such as one's ability to navigate the health care system [11]. Finally, by surveying patients immediately after they met with a clinician, our study demonstrates the value of gathering immediate perspectives of patients on their experiences with PGD sharing, which allowed us to quantitatively explore the interrelationships between patients' data tracking practices, data sharing practices, and their health literacy (Figure 1). In the future, similar approaches can be used to examine the immediate perspectives of clinicians on their experiences in responding to and using PGD shared by their patients.

Figure 1. Interrelationships of PGD tracking, PGD sharing, and health literacy (HSI and AE). AE: the ability to actively engage with health care providers; HSI: having sufficient information to manage one's health; PGD: patient-generated data; RQs: research questions.



Conclusions

We presented an onsite survey study of 109 patients who had just met with a clinician at a UHC. Our aim was to investigate the potential relationships between patients' PGD tracking practice, sharing practices, and their health literacy. We found that neither the number of data types the participants tracked nor how they tracked their data related to participants' perceptions that they had sufficient information to manage their health (HSI). However, participants who reported tracking more types of PGD provided higher ratings when asked about their ability to actively engage with their health care providers. Our results also highlighted that tracking PGD does not always lead to sharing these data with clinicians, as exemplified by the many barriers participants reported regarding data sharing. Reflecting on our findings, we discussed ways to close the gaps between patient tracking and sharing of their PGD. We also suggested design opportunities to improve patient health literacy by leveraging the value of PGD to support patients in assessing the relevance of their PGD to their health and to facilitate patient data sharing with their clinicians.

Acknowledgments

The authors would like to thank all the study participants for their time to share personal experiences. This research was in part funded by the National Science Foundation Awards (#1753452 and #1753453) and by the Research Improvement Grant from the College of Information Studies at the University of Maryland, College Park.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire.
[DOCX File , 19 KB - jmir_v22i12e18937_app1.docx]

References

1. Zhu H, Colgan J, Reddy M, Choe EK. Sharing patient-generated data in clinical practices: an interview study. AMIA Annu Symp Proc 2016;2016:1303-1312 [FREE Full text] [Medline: 28269928]
2. Wood W, Bennett AV, Basch E. Emerging uses of patient generated health data in clinical research. Mol Oncol 2015 May;9(5):1018-1024 [FREE Full text] [doi: 10.1016/j.molonc.2014.08.006] [Medline: 25248998]
3. Luo Y, Liu P, Choe EK. Co-designing Food Trackers With Dietitians: Identifying Design Opportunities for Food Tracker Customization. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; May 9-10, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300822]
4. Zhu H, Luo Y, Choe EK. Making Space for the Quality Care: Opportunities for Technology in Cognitive Behavioral Therapy for Insomnia. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI'17; April 3-7, 2017; Denver, CO, USA. [doi: 10.1145/3025453.3025549]
5. Schroeder J, Hoffswell J, Chung C, Fogarty J, Munson S, Zia J. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. CSCW Conf Comput Support Coop Work 2017 Feb 25;2017:1726-1739 [FREE Full text] [doi: 10.1145/2998181.2998276] [Medline: 28516172]
6. Chung C, Cook J, Bales E, Zia J, Munson SA. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. J Med Internet Res 2015 Aug 21;17(8):e203 [FREE Full text] [doi: 10.2196/jmir.4364] [Medline: 26297627]

7. Kim Y, Heo E, Lee H, Ji S, Choi J, Kim J, et al. Prescribing 10,000 Steps Like Aspirin: Designing a Novel Interface for Data-driven Medical Consultation. In: Conference on Human Factors in Computing Systems - Proceedings Internet. New York, NY. 2017 Presented at: CHI'17; April 7-10, 2017; Denver, CO, USA URL: https://dl.acm.org/doi/10.1145/3025453.3025570 [doi: 10.1145/3025453.3025570]
8. West P, Giordano R, Van KM, Shadbolt N. Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; May 5-7, 2016; San Jose, CA, USA URL: https://dl.acm.org/doi/10.1145/2858036.2858445 [doi: 10.1145/2858036.2858445]
9. Nutbeam D. Health promotion glossary. Health Promot 1986 May;1(1):113-127. [doi: 10.1093/heapro/1.1.113] [Medline: 10318625]
10. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
11. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the health literacy questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
12. Klasnja P, Kendall L, Pratt W, Blondon K. Long-term engagement with health-management technology: a dynamic process in diabetes. AMIA Annu Symp Proc 2015;2015:756-765 [FREE Full text] [Medline: 26958211]
13. Chung C, Dew K, Cole A, Zia J, Fogarty J, Kientz J, et al. Boundary Negotiating Artifacts in Personal Informatics: Patient-provider Collaboration With Patient-generated Data. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2016 Presented at: ACM'16; June 2-5, 2016; San Francisco, CA, USA. [doi: 10.1145/2818048.2819926]
14. Choe E, Lee N, Lee B, Pratt W, Kientz J. Understanding Quantified-selfers? Practices in Collecting and Exploring Personal Data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: SIGCHI'14; June 3-6, 2014; Toronto, Canada. [doi: 10.1145/2556288.2557372]
15. Choe EK, Lee B, Andersen TO, Wilcox L, Fitzpatrick G. Harnessing the power of patient-generated data. IEEE Pervasive Comput 2018 Apr;17(2):50-56. [doi: 10.1109/mprv.2018.022511243]
16. Choe EK, Lee B, Zhu H, Riche N. Understanding Self-reflection: How People Reflect on Personal Data Through Visual Data Exploration. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: EAI'17; June 1-5, 2017; Barcelona, Spain. [doi: 10.1145/3154862.3154881]
17. Choe EK, Lee B, Kay M, Pratt W, Kientz J. SleepTight: Low-burden, Self-monitoring Technology for Capturing and Reflecting on Sleep Behaviors. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: ACM'15; July 16-18, 2015; Osaka, Japan. [doi: 10.1145/2750858.2804266]
18. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 9;13:77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
19. Ishikawa H, Yano E. Patient health literacy and participation in the health-care process. Health Expect 2008 Jun;11(2):113-122 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00497.x] [Medline: 18494956]
20. 23 and Me. URL: https://www.23andme.com/ [accessed 2020-11-23]
21. Ayobi A, Marshall P, Cox A, Chen Y. Quantifying the Body and Caring for the Mind: Self-tracking in Multiple Sclerosis. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI'17; August 3-6, 2017; Denver, CO, USA. [doi: 10.1145/3025453.3025869]
22. Karkar R, Schroeder J, Epstein D, Pina L, Scofield J, Fogarty J, et al. TummyTrials: a Feasibility Study of Using Self-experimentation to Detect Individualized Food Triggers. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI'17; May 1-7, 2017; Denver, CO, USA. [doi: 10.1145/3025453.3025480]
23. Eikey EV. It's Definitely Been a Journey?: a Qualitative Study on How Women With Eating Disorders Use Weight Loss Apps. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI'17; May 1-7, 2017; Denver, CO, USA. [doi: 10.1145/3025453.3025591]
24. Eikey EV. Providers'perceptions of the Impact of Weight Loss Apps on Users With Eating Disorders. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; June 23-26, 2016; San Jose, CA, USA. [doi: 10.1145/2890602.2906194]
25. Garrity TF. Medical compliance and the clinician-patient relationship: a review. Soc Sci Med Part E: Med Psychol 1981 Aug;15(3):215-222. [doi: 10.1016/0271-5384(81)90016-8]
26. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013 Mar 27;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
27. Sands DZ, Wald JS. Transforming health care delivery through consumer engagement, health data transparency, and patient-generated health information. Yearb Med Inform 2018 Mar 5;23(1):170-176. [doi: 10.15265/iy-2014-0017]
28. West P, Van Kleek M, Giordano R, Weal M, Shadbolt N. Information quality challenges of patient-generated data in clinical practice. Front Public Health 2017;5:284 [FREE Full text] [doi: 10.3389/fpubh.2017.00284] [Medline: 29209601]

29. Jo DM. Issue Brief: Patient-Generated Health DataHealth IT. Office of the National Coordinator for Health Information Technology. 2013. URL: https://www.healthit.gov/sites/default/files/pghd_brief_final122013.pdf [accessed 2020-11-23]
30. Patient-Generated Health Information Technical Expert Panel. Department Services, of Health and Human. 2013. URL: https://www.healthit.gov/sites/default/files/pghi_tep_finalreport121713.pdf [accessed 2020-11-23]
31. Ostherr K, Borodina S, Bracken RC, Lotterman C, Storer E, Williams B. Trust and privacy in the context of user-generated health data. Big Data Soc 2017 Apr 17;4(1):205395171770467. [doi: 10.1177/2053951717704673]
32. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Informatics Assoc 2008 Jan 1;15(1):1-7. [doi: 10.1197/jamia.m2562]
33. Carmona RH. Health literacy: a national priority. J Gen Intern Med 2006 Aug;21(8):803. [doi: 10.1111/j.1525-1497.2006.00569.x]
34. Cho YI, Lee SD, Arozullah AM, Crittenden KS. Effects of health literacy on health status and health service utilization amongst the elderly. Soc Sci Med 2008 Apr;66(8):1809-1816. [doi: 10.1016/j.socscimed.2008.01.003] [Medline: 18295949]
35. Noblin A, Wan T, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag 2012;9:1-13 [FREE Full text] [Medline: 23209454]
36. St Leger L. Schools, health literacy and public health: possibilities and challenges. Health Promot Int 2001 Jun;16(2):197-205. [doi: 10.1093/heapro/16.2.197] [Medline: 11356758]
37. Williams MV, Davis T, Parker R, Weiss B. The role of health literacy in patient-physician communication. Fam Med 2002 May;34(5):383-389. [Medline: 12038721]
38. Ishikawa H, Yano E, Fujimori S, Kinoshita M, Yamanouchi T, Yoshikawa M, et al. Patient health literacy and patient-physician information exchange during a visit. Fam Pract 2009 Dec;26(6):517-523. [doi: 10.1093/fampra/cmp060] [Medline: 19812242]
39. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health Informatics J 2016 Jun;22(2):81-98 [FREE Full text] [doi: 10.1177/1460458214534092] [Medline: 24916567]
40. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults. J Gen Intern Med 1995 Oct;10(10):537-541. [doi: 10.1007/bf02640361]
41. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
42. Friis K, Lasgaard M, Osborne R, Maindal H. Gaps in understanding health and engagement with healthcare providers across common long-term conditions: a population survey of health literacy in 29,473 Danish citizens. BMJ Open 2016 Jan 14;6(1):e009627 [FREE Full text] [doi: 10.1136/bmjopen-2015-009627] [Medline: 26769783]
43. Hawkins M, Gill SD, Batterham R, Elsworth GR, Osborne RH. The Health Literacy Questionnaire (HLQ) at the patient-clinician interface: a qualitative study of what patients and clinicians mean by their HLQ scores. BMC Health Serv Res 2017 Apr 27;17(1):309 [FREE Full text] [doi: 10.1186/s12913-017-2254-8] [Medline: 28449680]
44. Cohen DJ, Keller SR, Hayes GR, Dorr DA, Ash JS, Sittig DF. Integrating patient-generated health data into clinical care settings or clinical decision-making: lessons learned from project healthdesign. JMIR Hum Factors 2016 Oct 19;3(2):e26 [FREE Full text] [doi: 10.2196/humanfactors.5919] [Medline: 27760726]
45. Nittas V, Lun P, Ehrler F, Puhan MA, Mütsch MA. Electronic Patient-Generated Health Data to Facilitate Disease Prevention and Health Promotion: Scoping Review. J Med Internet Res 2019 Oct 14;21(10):e13320 [FREE Full text] [doi: 10.2196/13320] [Medline: 31613225]
46. Qualtrics. 2019. URL: https://www.qualtrics.com/ [accessed 2020-11-23]
47. Porter K, Chen Y, Estabrooks P, Noel L, Bailey A, Zoellner J. Using teach-back to understand participant behavioral self-monitoring skills across health literacy level and behavioral condition. J Nutr Educ Behav 2016 Jan;48(1):20-6.e1 [FREE Full text] [doi: 10.1016/j.jneb.2015.08.012] [Medline: 26453368]
48. Zoellner JM, Hedrick VE, You W, Chen Y, Davy BM, Porter KJ, et al. Effects of a behavioral and health literacy intervention to reduce sugar-sweetened beverages: a randomized-controlled trial. Int J Behav Nutr Phys Act 2016;13():38. [doi: 10.1186/s12966-016-0362-1] [Medline: 27000402]
49. Vamos S, Klein K. Our self-tracking movement and health literacy: are we really making every moment count? Glob Health Promot 2018 Jun;25(2):85-89. [doi: 10.1177/1757975916660674] [Medline: 27488673]
50. MyFitnessPal. URL: https://www.myfitnesspal.com/ [accessed 2020-11-23]
51. Ferguson LA. Implementing a video education program to improve health literacy. J Nurse Pract 2012 Sep;8(8):e17-e22. [doi: 10.1016/j.nurpra.2012.07.025]
52. Yli-Kauhaluoma S, Pantzar M. Seeking connectivity to everyday health and wellness experiences: specificities and consequences of connective gaps in self-tracking data. Digit Health 2018;4:2055207618779714 [FREE Full text] [doi: 10.1177/2055207618779714] [Medline: 29942640]
53. Choe EK. Design of Persuasive Technologies for Healthy Sleep Behavior. In: Proceedings of the 13th international conference on Ubiquitous computing. 2011 Presented at: UbiComp'11; April 7-9, 2011; Beijing, China. [doi: 10.1145/2030112.2030193]
54. Liang Z, Alberto M, Martell C. Framing Self-quantification for Individual-level Preventive Health Care. In: Proceedings of the International Conference on Health Informatics. 2015 Presented at: CHI'15; May 7-9, 2015; Lisbon, Portugal. [doi: 10.5220/0005202503360343]

55. Verdezoto N, Grönvall E. On preventive blood pressure self-monitoring at home. Cogn Tech Work 2015 Nov 23;18(2):267-285. [doi: 10.1007/s10111-015-0358-7]
56. den Hoed M, Eijgelsheim M, Esko T, Brundel BJJM, Peal DS, Evans DM. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. Nat Genet 2013 Jun;45(6):621-631 [FREE Full text] [doi: 10.1038/ng.2610] [Medline: 23583979]
57. Fitbit. URL: https://www.fitbit.com/ [accessed 2020-11-23]
58. Choe EK, Abdullah S, Rabbi M, Thomaz E, Epstein DA, Cordeiro F, et al. Semi-automated tracking: a balanced approach for self-monitoring applications. IEEE Pervasive Comput 2017 Jan;16(1):74-84. [doi: 10.1109/mprv.2017.18]
59. Kim Y, Jeon JH, Lee B, Choe EK, Seo J. OmniTrack. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-28. [doi: 10.1145/3130930]
60. West P, Van KM, Giordano R, Weal M, Shadbolt N. Common Barriers to the Use of Patient-generated Data Across Clinical Settings. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; July 17-19, 2018; Montreal, Canada. [doi: 10.1145/3173574.3174058]
61. Use Fitbit's New Wellness Report To Have A More Informed Conversation With Your Doctor. URL: https://blog.fitbit.com/fitbit-premium-wellness-report/ [accessed 2020-11-23]
62. Lose It!. URL: https://www.loseit.com/ [accessed 2020-11-23]
63. Lupton D. 'I just want it to be done, done, done!' food tracking apps, affects, and agential capacities. MTI 2018 May 23;2(2):29. [doi: 10.3390/mti2020029]
64. Morris RL, Soh S, Hill KD, Buchbinder R, Lowthian JA, Redfern J, et al. Measurement properties of the Health Literacy Questionnaire (HLQ) among older adults who present to the emergency department after a fall: a Rasch analysis. BMC Health Serv Res 2017 Aug 29;17(1):605 [FREE Full text] [doi: 10.1186/s12913-017-2520-9] [Medline: 28851344]
65. Cutilli CC, Bennett IM. Understanding the health literacy of America. Orthopaedic Nursing 2009;28(1):27-32. [doi: 10.1097/01.nor.0000345852.22122.d6]
66. Debussche X, Lenclume V, Balcou-Debussche M, Alakian D, Sokolowsky C, Ballet D, et al. Characterisation of health literacy strengths and weaknesses among people at metabolic and cardiovascular risk: Validity testing of the Health Literacy Questionnaire. SAGE Open Med 2018;6:2050312118801250 [FREE Full text] [doi: 10.1177/2050312118801250] [Medline: 30319778]
67. Brabers AE, Rademakers JJ, Groenewegen PP, van Dijk L, de Jong JD. What role does health literacy play in patients' involvement in medical decision-making? PLoS One 2017;12(3):e0173316 [FREE Full text] [doi: 10.1371/journal.pone.0173316] [Medline: 28257472]
68. Figueiredo M, Caldeira C, Chen Y, Zheng K. Routine self-tracking of health: reasons, facilitating factors, and the potential impact on health management practices. AMIA Annu Symp Proc 2017;2017:706-714. [Medline: 29854136]

Abbreviations

AE: the ability to actively engage with health care providers
eHEALS: eHealth literacy scale
HLQ: health literacy questionnaire
HSI: having sufficient information to manage one’s health
IBS: irritable bowel syndrome
PGD: patient-generated data
RQ: research question
TOFHLA: Test of Functional Health Literacy in Adults
UHC: University Health Center

Edited by G Eysenbach; submitted 27.03.20; peer-reviewed by G Fitzpatrick, L Emmerton; comments to author 12.06.20; revised version received 07.08.20; accepted 26.10.20; published 22.12.20.

Please cite as:
Luo Y, Oh CY, Jean BS, Choe EK
Interrelationships Between Patients' Data Tracking Practices, Data Sharing Practices, and Health Literacy: Onsite Survey Study
J Med Internet Res 2020;22(12):e18937
URL: http://www.jmir.org/2020/12/e18937/
doi:10.2196/18937
PMID:33350960

Original Paper

Capitated Telehealth Coaching Hospital Readmission Service in Australia: Pragmatic Controlled Evaluation

Carmel Martin[1], MBBS, PhD, MSc, FRACGP, FAFPHM; Narelle Hinkley[2], BSc; Keith Stockman[2], BSc, MA; Donald Campbell[3], MBBS, MD, MMedSci, FRACP

[1]Monash Health Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Australia

[2]Community Health, Monash Health, Dandenong, Victoria, Australia

[3]Northern Health, Northern Hospital, Epping, Victoria, Australia

Corresponding Author:
Carmel Martin, MBBS, PhD, MSc, FRACGP, FAFPHM
Monash Health Faculty of Medicine
Nursing and Health Sciences
Monash University
Monash Medicine (Level 5)
35 Rainforest Walk
Clayton, VIC 3800
Australia
Phone: 61 3 9905 5035
Email: carmelmarymartin@gmail.com

Abstract

Background: MonashWatch is a telehealth public hospital outreach pilot service as a component of the Government of Victoria's statewide redesign initiative called HealthLinks: Chronic Care. Rather than only paying for hospitalizations, projected funding is released earlier to hospitals to allow them to reduce hospitalization costs. MonashWatch introduced a web-based app, Patient Journey Record System, to assess the risk of the journeys of a cohort of patients identified as frequent admitters. Telecare guides call patients using the Patient Journey Record System to flag potential deterioration. Health coaches (nursing and allied health staff) triage risk and adapt care for individuals.

Objective: The aim was a pragmatic controlled evaluation of the impact of MonashWatch on the primary outcome of bed days for acute nonsurgical admissions in the intention-to-treat group versus the usual care group. The secondary outcome was hospital admission rates. The net promoter score was used to gauge satisfaction.

Methods: Patients were recruited into an intention-to-treat group, which included active telehealth and declined/lost/died groups, versus a systematically sampled (4:1) usual care group. A rolling sample of 250-300 active telehealth patients was maintained from December 23, 2016 to June 23, 2019. The outcome—mean bed days in intervention versus control—was adjusted using analysis of covariance for age, gender, admission type, and effective days active in MonashWatch. Time-series analysis tested for trends in change patterns.

Results: MonashWatch recruited 1373 suitable patients who were allocated into the groups: usual care (n=293) and intention-to-treat (n=1080; active telehealth: 471/1080, 43.6%; declined: 485, 44.9%; lost to follow-up: 178 /1080, 10.7%; died: 8/1080, 0.7%). Admission frequency of intention-to-treat compared to that of the usual care group did not significantly improve (*P*=.05), with a small number of very frequent admitters in the intention-to-treat group. Age, MonashWatch effective days active, and treatment group independently predicted bed days. The analysis of covariance demonstrated a reduction in bed days of 1.14 (*P*<.001) in the intention-to-treat group compared with that in the usual care group, with 1236 bed days estimated savings. Both groups demonstrated regression-to-the-mean. The downward trend in improved bed days was significantly greater (*P*<.001) in the intention-to-treat group (Sen slope –406) than in the usual care group (Sen slope –104). The net promoter score was 95% in the active telehealth group compared with typical hospital scores of 77%.

Conclusions: Clinically and statistically meaningful reductions in acute hospital bed days in the intention-to-treat group when compared to that of the usual care group were demonstrated (*P*<.001), although admission frequency was unchanged with more short stay admissions in the intention-to-treat group. Nonrandomized control selection was a limitation. Nonetheless, MonashWatch was successful in the context of the HealthLinks: Chronic Care capitation initiative and is expanding.

(J Med Internet Res 2020;22(12):e18046) doi:10.2196/18046

KEYWORDS

telehealth; funding model; evaluation; health services research; potentially preventable hospitalizations, medical informatics

Introduction

Overview

Potentially preventable hospitalizations or potentially avoidable admission costs are of significant interest, not only to governments and hospitals, but to individuals, their families, the community, and general practice [1]. A pragmatic study evaluated the impact of MonashWatch, a telehealth coaching capitated pilot service in Victoria, Australia, on bed days in the context of a statewide rollout of a new funding model. Rather than only paying for hospitalizations, projected admission funding is released in advance to hospitals to allow them to develop systems that will reduce preventable hospitalizations.

HealthLinks: Chronic Care (HLCC) is a voluntary, funding-neutral reform that aims to support the Australian State of Victoria's public health services in adopting outcome-based, rather than activity-based, funding [2]. An algorithm running on hospital data identifies patients at-risk of potentially preventable hospitalizations and informs participating hospitals who can use financing from anticipated admissions to address care needs better and earlier.

The MonashWatch telehealth and coaching model used design principles to establish a collaborative patient-journey approach responding to broad social determinants beyond disease management and the boundaries of hospital, primary, home, and social care.

Laypersons called *telecare guides* track risk and identify issues in biopsychosocial and environmental domains using frequent telephone calls and the Patient Journey Record System, which uses a client-server architecture with a browser-based user interface. A rule-based algorithm provides a real-time risk assessment of calls based on data entry and telecare guide opinion. *Health coaches* triage calls and support participants to optimize their health journeys.

This paper reports a pragmatic summative evaluation of the MonashWatch service. We compared bed days for an intention-to-treat group versus a usual care control group for 30 months from the MonashWatch service commencement. The intention-to-treat group included a MonashWatch active telehealth group consisting of those who used the telehealth service.

Background

The Australia-wide universal free public health system has both federal and state/territorial governance. Parallel private health systems exist. Medicare is the Australian federal government's scheme to give universal public access to health care (funded by taxation—the Medicare levy) through (1) direct clinical service funding to general practitioners and specialists in all states and territories and (2) indirect financing, with the states and territories administering public hospital and most community services. Most services, including social services and welfare, aged care, education, and employment, have split funding and administration across federal and state/territory systems.

There have been multiple initiatives to address avoidable admission costs across the jurisdictions for more than 20 years. The Australian Institute of Health and Welfare (AIHW) annually reports on 18 International Statistical Classification of Diseases, Tenth Revision (ICD-10) diagnostic codes of chronic, acute, and vaccination-preventable admissions for national performance monitoring by local area [1]. Since 2000, there have been two significant waves of coordinated care trials which pooled federal and state resources (and sometimes private health resources) in several local trials to improve potentially preventable hospitalizations costs and health outcomes [3]. Significantly higher health service use and costs were incurred in the absence of clear evidence of improved health outcomes. Many clients did not require care coordination. Funds pooling arrangements contributed to limited possibilities for service substitution, and training of general practitioner care coordinators was inadequate [3]. Ambitious large-scale randomized controlled methods in health service transformation trials, which have failed due to the poor implementation of live services, have been part of the problem [3]. Ongoing and continually changing federal and state/territory integrated care initiatives have similarly failed to document improvements for care costs and outcomes [4]. Nevertheless, the Australian health system is high performing. In 2016, health expenditure as the proportion of gross domestic product was 9.6% for Australia, 9.0% for all Organization for Economic Co-operation and Development countries, 9.7% for the United Kingdom, and 10.6% for Canada; and Australia's healthy life expectancy is 73 years, around 10 years higher than the global average life expectancy [5].

Victorian Government Funding Reform Initiative in Acute Hospital Use

Victoria is the second most populous state of Australia with a population of 6.25 million. The Victorian Department of Health and Human Services (DHHS) funds and administers 85 public health service organizations. It has had activity-based funding (fee-for-service) for all acute admissions with accelerating demand growth in hospital separations per 1000 individuals. Victoria has deployed hospital readmission prevention programs since 1996, which have improved satisfaction with care but have had little impact on cost containment [6].

The HLCC initiative began in 2016, identifying patients at-risk of ≥3 repeat hospitalizations and their admission costs in the subsequent 12 months [7]. The DHHS incentivizes participating hospital systems to improve admission bed days, costs, and care quality within projected costs (HLCC-identified patients).

Monash Health

Monash Health is the most extensive public hospital and community care system in Victoria. Its 15,000 staff (with a large hospital readmission prevention program base) work at more than 40 sites, providing more than 3 million occasions of service, admitting more than 238,000 hospital patients, and handling more than 206,000 emergency presentations per year. HLCC data indicated that in 2017 and 2018, Monash Health had more than 3000 patients with >4 acute medical admissions and more than 12,000 with >3 acute medical admissions (30% of which were potentially preventable hospitalizations). It was an early adopter of the HLCC service initiative MonashWatch in 2016.

Rather than adding a new layer, the MonashWatch model was intended to be a catalyst within a working health system, to begin a transition of acute services to outcome-based funding.

MonashWatch Model of Care

Patient journeys involve physiological, psychological, social, and environmental issues. Disturbances in any or combinations of domains including housing, food security, support with daily living, access issues, as well as biology, medication issues, or clinical deterioration may lead to tipping points into acute admissions [8,9]. Therefore, MonashWatch focused on a journey model to enable individuals to optimize their health trajectory with their caregivers and essential people, rather than focus mainly on their selected diseases and treatment adherence [10]. The Patient Journey Record System was designed to monitor health journeys and was initially developed in Ireland and validated in an Irish primary care cohort [11,12]. A key feature is that telecare guides, who are nonprofessionals, engage with people from their community to monitor and support their care. The structure of the Patient Journey Record online system and the supervision of coaches ensure the quality and safety of telecare guide work.

Clinicians—nursing or allied health—enable others in the goals and health journeys which they chose to follow [13]. Coaching, in the MonashWatch context, is an innovative flexible transdisciplinary role. The MonashWatch coaching role, incorporating the Patient Journey Record System journey model, empowers clinicians from multiple disciplines to bring their specific expertise and collaborate across disciplinary boundaries to enable person-centered rather than professional-centered siloed care. The role was developed through extensive consultation with patients and advocacy groups, service providers, and the DHHS. The role of the coach providing oversight and support for telecare guides, as well as individual patient coaching, is essential. The advantage of a multidisciplinary team role is the very broad scope of integrated practice—nursing, physiotherapy, social work, occupational therapy, and medicine. A coaching framework was constructed to address health, resilience, and need perceptions rather than protocol-driven care and to guide people through a health and welfare maze. Coaches use HLCC-capitated funds for food, transport to outpatients, or general practitioners, or a second opinion in the private sector as needed.

Active Telehealth Service

Following HLCC algorithm identification and allocation, the intention-to-treat group were invited to initially consent and agree to a home visit for enrollment, formal consent, baseline assessment, and induction. Telecare guides, then made conversational phone calls to enrolled people to track their health and needs, in accordance with personal preference and previous Patient Journey Record System flags.

Regular audiotaped calls between 1 to 5 times per week (median 1), depending on risk level, were conducted by telecare guides. They used the Patient Journey Record System semistructured monitoring app, which began with open-ended narratives and included directed questions [11] in a branching format (Figure 1) Patient Journey Record System flags were traffic-light indicators of physical and psychosocial resilience and symptomatology generated from the conversations in real time using an internal algorithm [11,12]. Flags implicated risk fluctuations in short-term intervals (hours to days) in individual journeys that may foreshadow potentially preventable hospitalizations admissions. Patient Journey Record System risk analysis guided telecare guide decision making on the timeframe (eg, immediately, tomorrow, etc) of the need to involve their health coaches. Telecare guides and coaches called the same individuals and got to know them over time, forming relationships. Coaches triaged calls and intervened to address urgent or nonurgent and high, medium, and lower risk issues promptly, including the needs of carers. Coaches worked directly with the general practitioner who was the primary medical provider. Coaches also worked directly with emergency departments; inpatients; outpatients; and drug, alcohol, and social services on an as-needed basis. Coaches triaged and anticipated risk of health deterioration using the PaJR predictive algorithm and human sensemaking [14]. They, then, identified and intervened in the root causes of readmissions, where possible [14] (see also Figure 2). Coaches triaged, navigated, and supported MonashWatch participants in a reactive and anticipatory manner bridging the gaps with general practitioner, hospital, pharmacy, social welfare, housing, legal, and other support services. Coaches made telephone contact, home visits, and accompanied participants to general practitioners, clinics, and other facilities, where appropriate.

Figure 1. Patient Journey Record semistructured monitoring app that begins with open-ended narratives and includes direct questions.

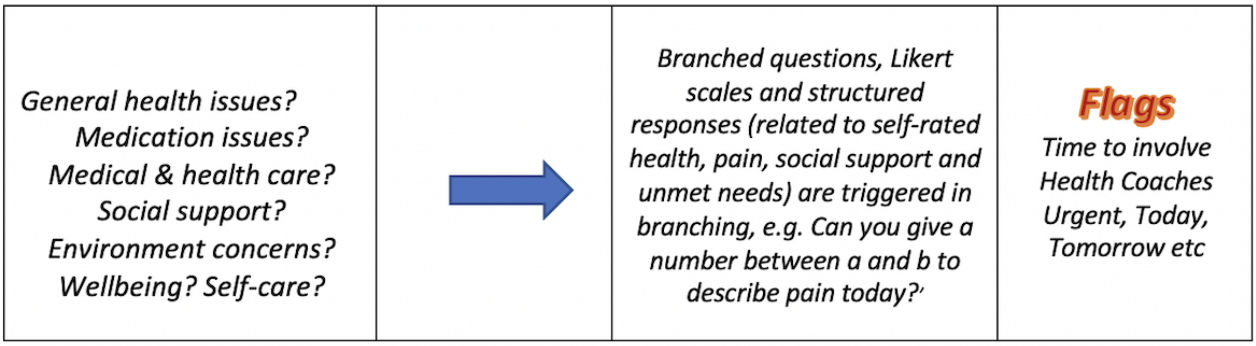
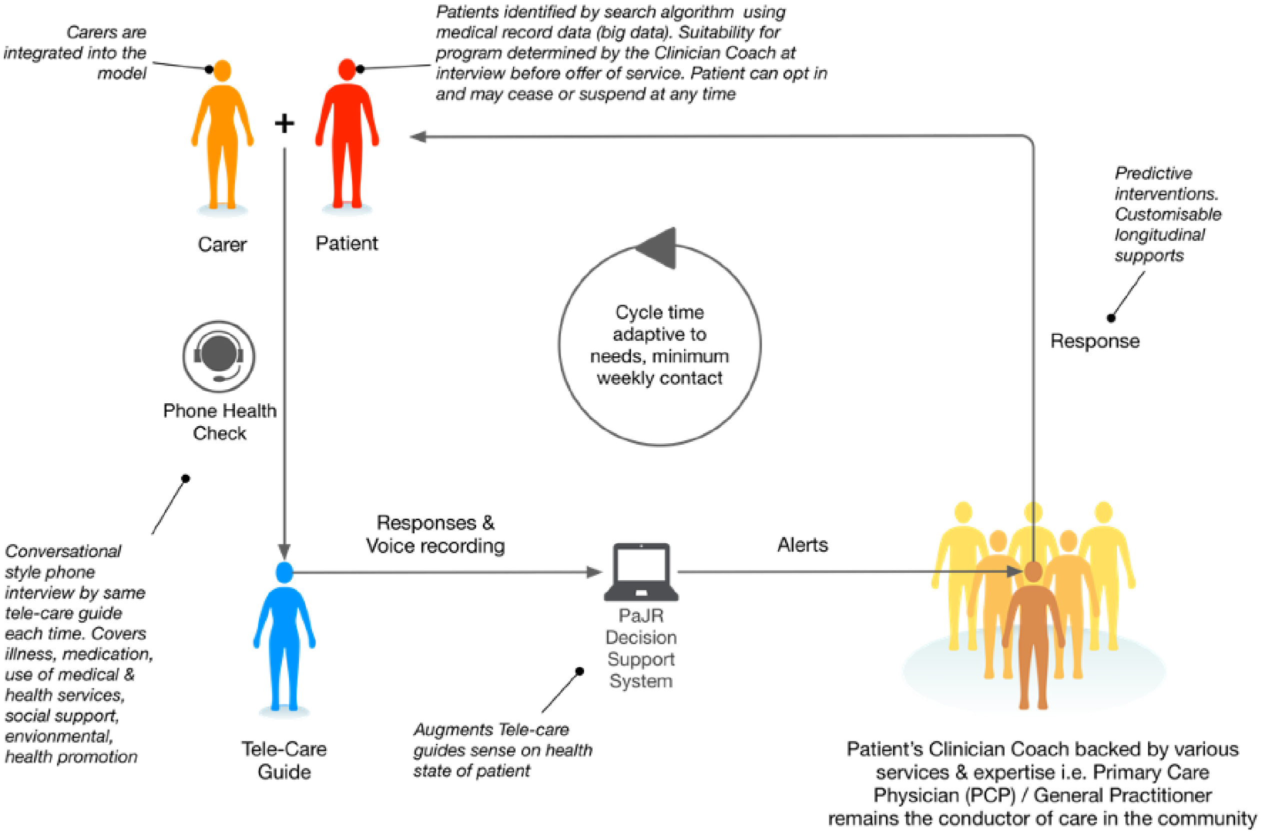


Figure 2. The MonashWatch model (reproduced from Martin et al [14]). PaJR: Patient Journey Record.



Methods

Implementation of the MonashWatch Model

The health coaches were all recruited from within the Monash Health community and acute services to develop the new service. Telecare guides were recruited from the local or adjacent community. Care pathways to existing services were mapped and tested before the service commenced. The pathways were to the general practitioners, hospitals' emergency, inpatient and outpatients, hospital readmission prevention programs, and other community health, social, and welfare services including housing, legal, financial, employment, education, and voluntary organizations.

Participants

MonashWatch participants were selected in a highly disadvantaged catchment area adjacent to the Dandenong hospital. Keeping the MonashWatch team local to patients was a concept borrowed from Buurtzorg: the Dutch neighborhood care model [15]. Targeted patients resided within a short car trip from the MonashWatch team location to minimize clinician travel time and hence maximize responsiveness and clinical time. Most admissions for patients in the cohort occurred at the local public secondary care hospital in Dandenong. However, some took place in other more distant Monash Health hospitals, subject primarily to demand and clinical needs.

Potentially Preventable Hospitalizations

The HLCC web-based algorithm incorporated a wide range of conditions in adults >18 years old [2]. The HLCC algorithm incorporated the following: (1) service parameters including rates of recent acute admissions and emergency department visits; (2) patient parameters including age, residence status, smoking; (3) chronic conditions such as gastrointestinal disorders, renal disease, asthma, chronic obstructive pulmonary disease, rheumatoid arthritis, diabetes, pancreatic conditions, cirrhosis/alcoholic hepatitis; but excluded serious mental and psychotic illnesses, dialysis, and cancer treatments because there are other initiatives for these groups.

AIHW potentially preventable hospitalizations diagnostic codes only accounted for 18% of HLCC admissions. Only approximately 20% of HLCC-identified hospitalizations in Monash Health, and other Victorian health services had ever accessed hospital readmission prevention programs or other hospital admission prevention services.

Participant Enrollment Allocation

The DHHS provides continuously updated HLCC-eligible cohort lists to hospital groups and funds care improvement initiatives based on projected reductions in admission costs. Once the patient was deemed eligible, when they had their next acute admission, they could be enrolled. Enrollment commenced with a gradual ramp-up from December 2016 and continued beyond the evaluation cutoff point. Pragmatic screening by health coaches excluded those who were not suited to a self-rated phone-based health model (eg, nursing home, necessitated use of an interpreter, and patients who would pose a high risk to staff visiting at home). Patients were considered candidates to be entered into the MonashWatch evaluation pool before allocation, based on a ratio of 4:1. There was minimal chance of bias because the health coaches and team performing the assignment had no idea who would benefit in advance in this pilot service, and the allocation was conducted using hospital unit numbers from a list without patient details.

Outcomes

The primary outcome metric was bed days (ie, length of stay related to emergency nonsurgical admissions) derived from the Victorian Admitted Episode Data from the Victorian Emergency Minimum Dataset [16] between December 23, 2016 and June 23, 2019.

A secondary outcome metric was rate of emergency nonsurgical admissions. This was initially considered as the primary outcome; however, capitation costs being the biggest driver of the HLCC program led to bed days being more critical. Net promoter score was also a secondary outcome.

Statistical and Other Methods

Analysis of covariance (ANCOVA) is a statistical technique that adjusts for covariates in determining the outcomes of an intervention. Least square means are an acceptable method to calculate the means adjusted for covariates [17]. In ANCOVA, least square means are group-mean adjusted for covariates (ie, holding constant at some typical value of the covariate, such as its mean value). Effective days active describes the duration in days post assignment to intention-to-treat or usual care. The outcome was least square of effective days active of control versus intervention, mean-adjusted for the quantitative variables above. Two-tailed Mann-Kendall trend test and Sen slope were used for the significance of the time series trend and to calculate an estimate of the trend. The net promoter score survey was anonymously administered (as a postal survey) with open-ended comments. The score was calculated following the Australian National Safety and Quality Health Service Standard [18]. The net promoter score is an index ranging from –100 to 100 that measures the willingness of customers to recommend a company's products or services to others. It is calculated as the difference between the percentage of promoters and detractors and is used as a proxy for gauging the patient’s overall satisfaction with a hospital’s product or service and the patient’s loyalty to the service [18]. Other secondary outcomes measures including baseline and sequential measures have previously been described [14]. A significance level of $P<.05$ was used. Analysis of bed days was conducted using least square ANCOVA, in accordance with acceptable practice [19]. Mean bed day values were adjusted for age, gender, time in the MonashWatch intervention (effective days active), and the presence of potentially preventable hospitalization—18 selected ICD-10 diagnostic codes—covariates [1].

Quality assurance of the data analysis was performed in several ways. One author conducted the ANCOVA using XLSTAT software (version 2020:4.1; Addinsoft). Two other authors independently analyzed the same dataset, after rechecking the download from Victorian Admitted Episode Data/Victorian Emergency Minimum Dataset [16], using R (version 3.5.2; R Studio). External evaluation was carried out on bed days and satisfaction with external controls using propensity-scoring rather than contemporaneous local controls and has not yet been formally reported.

Ethics approval for low-risk clinical research was obtained from the Monash Health's Health Research Ethics Committee. The Australian government’s main research and development agency, the Commonwealth Scientific and Industrial Research Organisation, is conducting an external evaluation of the diverse state-wide HLCC initiatives in Victoria, and this also has ethics approval.

Results

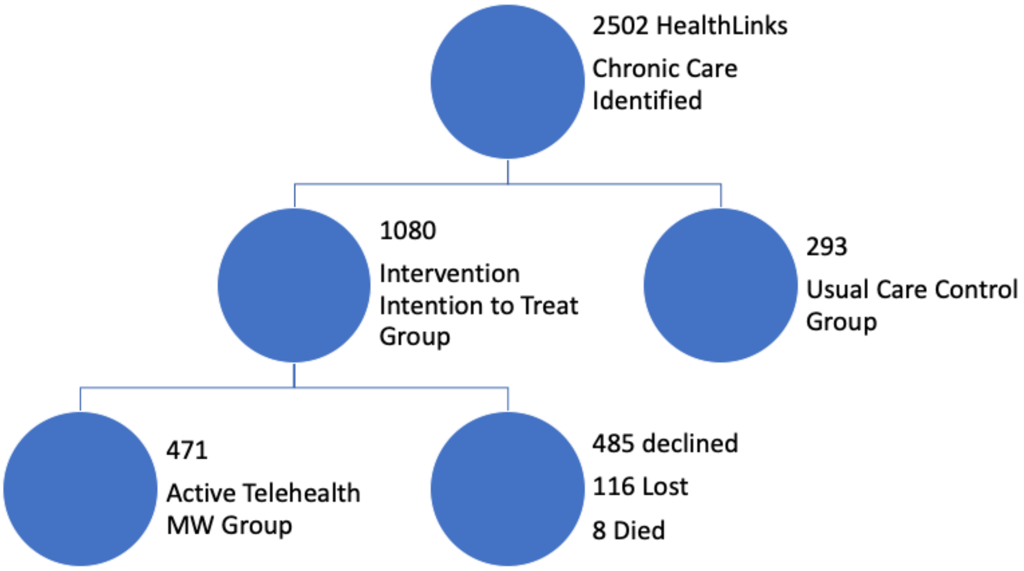
Implementation of the MonashWatch Model

The HLCC clinical algorithm identified 2502 patients as having a high risk of repeat admissions within the period of December 23, 2016 to June 23, 2019.

Participants

MonashWatch identified 1373 suitable HLCC patients: usual care (n=293) and intention-to-treat (all: n=1080; active telehealth: 471/1080, 43.6%; declined: 485/1080, 44.9%; lost to follow-up: 116/1080, 10.7%; died: 8/1080, 0.7%; Figure 3). The intention-to-treat active telehealth group provided consent to being recruited, and general practitioners, health, and social services were aware of their pilot service status—blinding was not possible. Controls had no contact with the MonashWatch team, and general practitioners, health, and social services were unaware that they were controls. Once allocated to the intention-to-treat and the active telehealth groups, patients remained in that group for the duration.

Figure 3. MonashWatch pragmatic clinical evaluation participants. Intention-to-treat, including active telehealth, and usual care cohort allocation in the MonashWatch pragmatic evaluation. MW: MonashWatch.



Mean participant age for usual care was 64.3 (SD 17.6; median 71, IQR 19) years, and mean participant age for intention-to-treat was 68.3 (SD 16.8; median 71, IQR 19) years. The number of effective days active for usual care was 756.2 (SD 180.5; median 1003, IQR 44), and the number of effective days active for intention-to-treat was 624.2 (SD 269.2; median 908, IQR 278).

Admissions

In the usual care group, 293 patients had 639 admissions, and in the intention-to-treat group, 1080 had 934 admissions from the time they joined the MonashWatch program (effective days active) until June 23, 2019 (Table 1). Usual care had 163/293 (55.6%) patients admitted at least once; a median of 3 admissions/person who was admitted (third quartile 5); a mean of 3.9 admissions/person who was admitted (SD 4.3, skewness 3.1). Intention-to-treat had 549 (50%) of patients admitted at least once; a median of 2 admissions/person who was admitted (third quartile 4); a mean of 3.6 admissions/person who was admitted (SD 4.3, skewness 6.8). Men were 44.2% (72/163) and women were 55.8% (91/163) of admissions in usual care. Men were 55.9% (307/549) and women were 44.1% (242/549) of admissions in intention-to-treat.

Table 1. Descriptive statistics on admission rates in MonashWatch intention-to-treat and usual care.

Statistic	Usual care	Intervention	*P* value
Admitted at least once, n (%)	163 (55.6)	549 (50.8)	.054
Description of profiles of those admitted in each group			
Median (third quartile)	3 (5)	2 (4)	.05
Mean (sample SD)	3.9 (4.3)	3.6 (4.6)	.056
Pearson skewness	3.1	6.8	<.001

Admissions were highly skewed in the intervention group with several outliers with frequent short admissions to the emergency department for chest pain and abdominal symptoms which may account for the highly skewed profile of the intention-to-treat group. The raw median and mean number of admissions per person were higher in the control group (*P*=.05 and *P*=.056, respectively; see Table 1).

Potentially Preventable Hospital Admissions

AIHW potentially preventable hospital admission codes were present in 18.3 % (117/639) of all acute admissions in usual care and 16.4% (153/ 934) of all acute admissions in intention-to-treat (*P*=.34), that is, 0.05% (8/1573) of all admissions were coded vaccine preventable. The most frequent AIHW potentially preventable hospitalizations conditions were chronic obstructive pulmonary disease or bronchiectasis (usual care: 30/641, 4.7%; intention-to-treat: 49/934, 5.2%; *P*=.19); heart failure or angina (usual care 17/641, 2.7%; intention-to-treat 24/934, 2.6%), urinary tract infection or cellulitis (usual care 32/642, 5.6%; intention-to-treat 30/934, 3.2%; *P*=.04). Chest pain (of minor complexity) was the most frequent admission description in all groups (Table 2). Two individuals, who called an ambulance for chest pain (of minor complexity) at least weekly and had frequent <1-day overnight stay admissions to the emergency department, were outliers in the MonashWatch active telehealth group. The nonactive telehealth group had fewer diagnoses of chest pain (of minor complexity), abdominal pain and mesenteric adenitis, minor complexity and other digestive system disorders, and minor complexity diagnoses than the intention-to-treat group. Low complexity conditions including chest pain and gastrointestinal conditions were coded (usual care: 86/641, 20.4%;

intention-to-treat: 219/1784, 12.2%). The usual care group had significantly more minor complex conditions diagnoses than the intention-to-treat group (*P*=.03). Diagnoses using AIHW potentially preventable hospitalization ICD codes that were designated as major complexity (high cost and high comorbidity) [20] are indicated. In the top 10 most frequent potentially preventable hospitalizations (Table 2), the usual care group had 3.0% (19/641) of admissions designated as major complexity, MonashWatch intention-to-treat active telehealth participants had 5.1% (48/934) of diagnoses identified as major complexity—significantly more than usual care (*P*=.02), and intention-to-treat nontelehealth participants had 1.1% (9/852) (*P*<.001).

Table 2. Most common diagnoses for potentially preventable hospitalizations as defined by the AIHW within each group of MonashWatch patients.

Group and diagnosis	Admission, n (%)
Control (usual care, n=641)	
Chest pain, minor complexity	46 (7.2)
Abdominal pain and mesenteric adenitis, minor complexity	20 (3.1)
Other digestive system disorders, major complexity	19 (3.0)
Bronchitis and asthma, minor complexity	18 (2.8)
Esophagitis and gastroenteritis, minor complexity	18 (2.8)
Arrhythmia, cardiac arrest and conduction disorders, minor complexity	16 (2.5)
Other digestive system disorders, minor complexity	16 (2.5)
Kidney and urinary tract infections, minor complexity	15 (2.3)
Syncope and collapse, minor complexity	15 (2.3)
Chronic obstructive airways disease, minor complexity	14 (2.2)
Intention-to-treat including active telehealth (n=934)	
Chest pain, minor complexity	110 (11.8)
Abdominal pain and mesenteric adenitis, minor complexity	41 (4.4)
Respiratory infections and inflammations (major complexity	29 (3.1)
Chronic obstructive airways disease, minor complexity	25 (2.7)
Syncope and collapse, minor complexity	22 (2.4)
Other digestive system disorders, minor complexity	21 (2.2)
Other digestive system disorders, major complexity	19 (2.0)
Headaches, minor complexity	18 (1.9)
Esophagitis and gastroenteritis, minor complexity	18 (1.9)
Arrhythmia, cardiac arrest and conduction disorders, minor complexity	16 (1.7)
Intervention (intention-to-treat) nontelehealth (n=390)	
Chest pain, minor complexity	41 (10.5)
Diabetes, minor complexity	10 (2.6)
Arrhythmia, cardiac arrest and conduction disorders, minor complexity	9 (2.3)
Abdominal pain and mesenteric adenitis, minor complexity	6 (1.5)
Chronic obstructive airways disease, minor complexity	6 (1.5)
Coronary atherosclerosis, minor complexity	6 (1.5)
Poisoning/toxic effects of drugs and other substances, minor complexity	6 (1.5)
Chronic obstructive airways disease, major complexity	5 (1.3)
Syncope and collapse, minor complexity	5 (1.3)
Heart failure and shock, major complexity	4 (1.0)

Primary Outcome

ANCOVA was conducted on bed days with admission age, gender, the presence or absence of a potentially preventable hospitalizations ICD-10 code, and effective days active as quantitative variables and with intervention versus control as qualitative variables

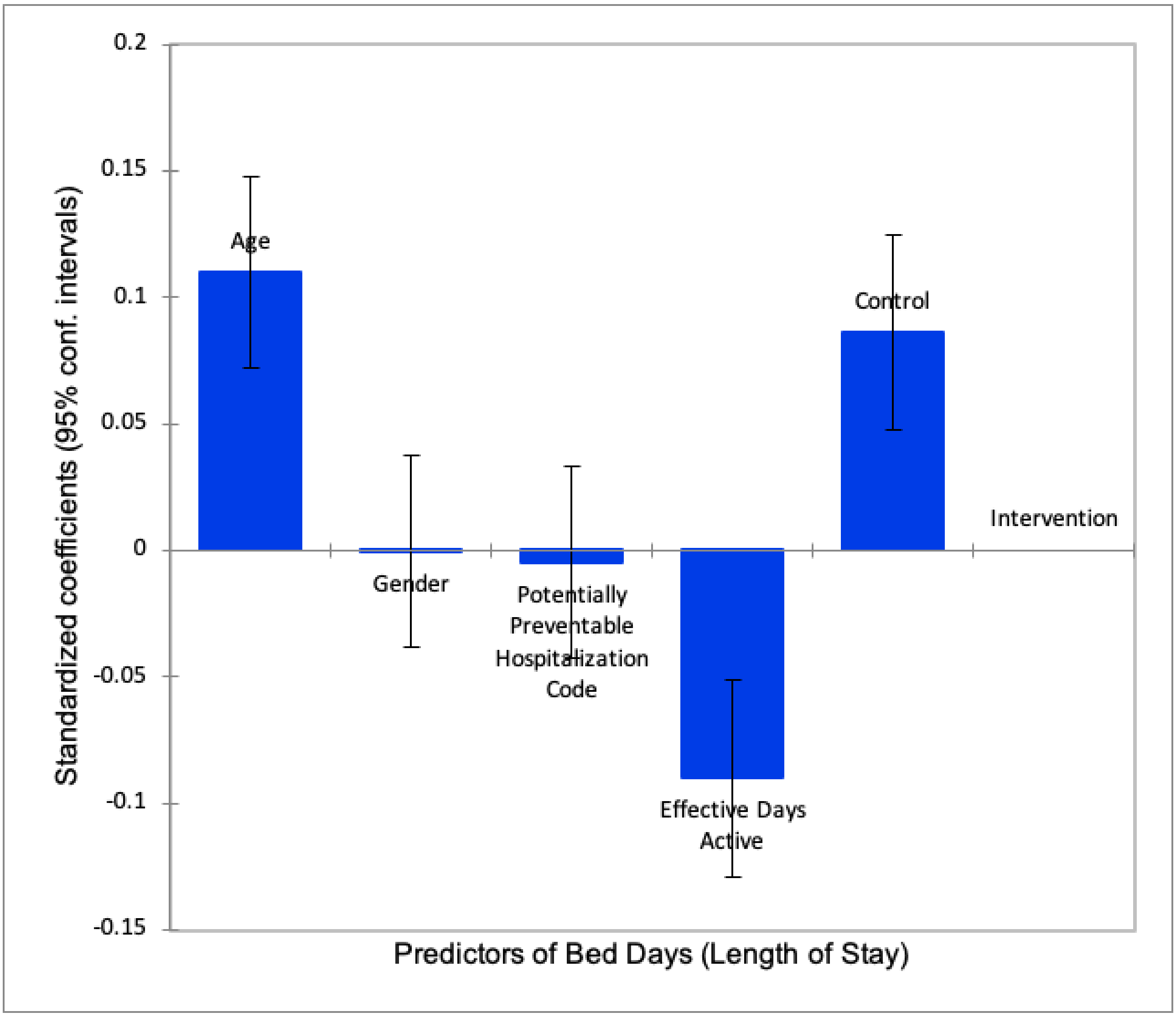
Table 3 and Figure 4 demonstrate the factors which significantly impacted on bed days in the sample: age (*P*<.001), effective days active from allocation until June 23, 2019 (*P*<.001), and whether they were in the intervention (intention-to-treat) or control (usual care) group (*P*<.001). Gender (*P*=.98) and having a designated AIHW potentially preventable hospitalizations admission ICD code (*P*=.82) were not significant predictors of bed days.

Table 3. ANCOVA summary statistics of the impact of key variables on bed days of MonashWatch patients: standardized coefficients predicting length of stay.

Source	Value	SE	*t* value	*P* value	95% CI
Admission age	0.110	0.019	5.673	<.001	(0.072, 0.148)
Gender	–0.001	0.019	–0.031	.98	(–0.038, 0.037)
AIHW[a] potentially preventable hospitalizations (0, no; 1, yes)	–0.005	0.019	–0.235	.82	(–0.042, 0.033)
Effective days active	–0.090	0.020	–4.566	<.001	(–0.129, –0.051)
Control (usual care) vs intervention (intention-to-treat)	0.086	0.020	4.348	<.001	(0.047, 0.125)

[a] AIHW: Australian Institute of Health and Welfare.

Figure 4. Bed days standardized coefficients based upon ANCOVA.



Age, MonashWatch effective days active, and intention-to-treat group status predicted bed days. The usual care least square mean was 4.5 (SD 0.2, 95% CI 4.1-4.9) bed days while the intention-to-treat least square mean was 3.4 (SD 0.1, 95% CI 3.1-3.6) bed days. A statistically significant (*P*<.001) bed days saving of 1.14 bed days per 1080 intention-to-treat patients (1236 days) was estimated (Table 4).

Table 4. ANCOVA least square mean bed days comparison for usual care versus intention-to-treat groups.

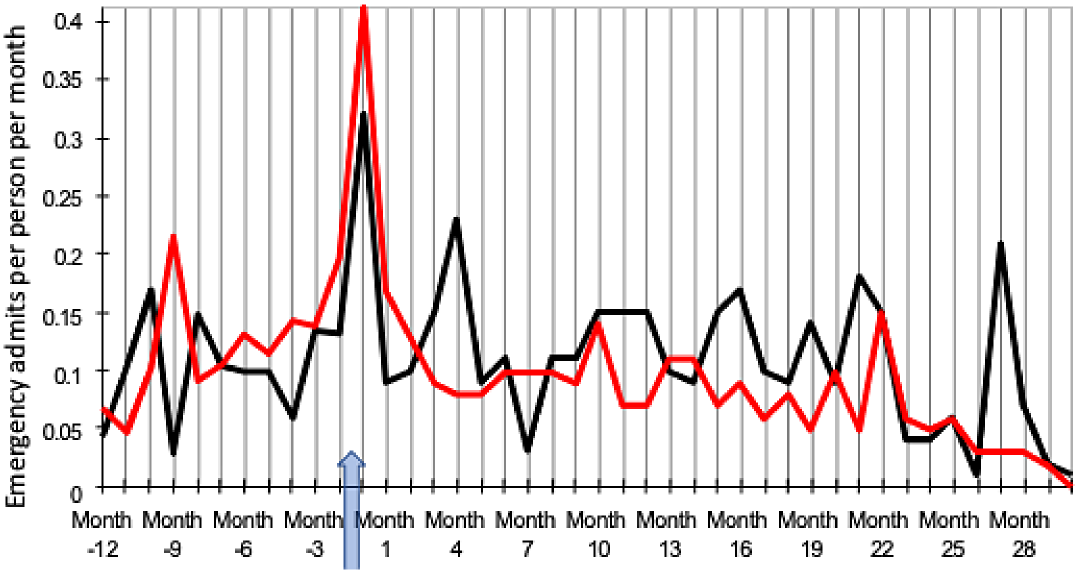
Category	Mean bed days	SE	95% CI	*P* value
Usual care	4.5	0.2	(4.1, 4.9)	<.001
Intention-to-treat	3.4	0.1	(3.1, 3.6)	

Usual Care Versus Intention-to-Treat

Longitudinal tracking of average bed days per person per month nonsurgical acute admissions was conducted for 12 months before enrollment until the evaluation cutoff date (Figure 5). The Mann-Kendall trend and Sen slope for intention-to-treat versus usual care demonstrated a significantly different trend in the intention-to-treat versus usual care time series. Rates of bed days per person per month in the 12 months before MonashWatch were higher in the intention-to-treat (1.77 bed days per person per month) versus usual care (1.44 bed days per person per month). Overall, intention-to-treat demonstrated statistically significant greater improvement (*P*<.001) in bed days per person per month (Kendall tau 0, sample variance 8483) compared with usual care (Kendall tau 0, sample variance 8458). The bed days per person per month time series demonstrated regression to the mean in both intention-to-treat and usual care groups. The intention-to-treat group demonstrated a statistically significant greater (*P*<.001) downward trend in bed days compared with the usual care group using Mann-Kendall trend test, and the Sen slope was –406 for intention-to-treat and the slope was –104 for usual care.

A high net-promoter rate (satisfaction scores) of 95% was demonstrated, with common findings of about 77% in hospital evaluations [18].

Figure 5. Rate of mean bed days per person per month in the usual care group (black) and intention-to-treat (red) 12 months prior to MonashWatch enrolment (indicated by the blue arrow) and for the subsequent 30 months. (Total admissions adjusted for numbers in each group.).



Discussion

Principal Findings

This evaluation demonstrates that the MonashWatch service intervention achieved its objectives by reducing bed days and, by implication, worked within the capitated budget consistently over time.

The percentage of patients admitted at least once was very similar (usual care: 55.6%; intention-to-treat: 50.8%). Admissions per person were nearly statistically significantly different (median, *P*=.05; mean, *P*=.056) with a frequent admitter intention-to-treat subgroup. The number of low complexity conditions, including chest pain and gastrointestinal conditions, was higher in the usual care group than it was in the intention-to-treat group. The intention-to-treat group had nonsignificantly more potentially preventable hospitalizations admissions, perhaps suggesting an optimization of hospital bed use, reducing admissions that were not related to clinical need alone. Analysis of covariance, controlling for age, effective days active, gender, and AIHW potentially preventable hospitalizations diagnoses demonstrated that bed days were statistically improved by 1.14 bed days per 1080 intention-to-treat patients (*P*<.001) and an estimated 1236 bed days. Improved bed days were consistent over the evaluation period. The intention-to-treat group demonstrated a consistent improvement in bed days per person per month (*P*<.001).

Challenges to constrain costs while improving care are prominent in the fragmented governance and funding system of Australia. Macrolevel federal and state reforms with pooled funding have not previously proven successful. Victoria has deployed state-based local initiatives to improve hospital readmissions since before 1996 but had identified the need to further “shift the dial [3,5,6].” A capitated funding model for hospital services to use projected admission costs earlier in a high admitting cohort was introduced in 2016 [2]. Monash Health took up this opportunity with unique internet-enabled telehealth (telephone) and coaching approach with real-time data usage. Using principles of disruptive innovation [21], MonashWatch formed an innovation hub within existing services and is becoming increasingly embedded in routine service delivery through its expansion and integration within other community services. Ongoing data-driven continuous improvement based on frequent audits and adaptation in the complex changing world of patient profiles and service changes will be a crucial feature [2].

Did the Study Findings Demonstrate Causal Associations?

The tracking of bed days and triangulation with other findings indicated a causal association between MonashWatch and the improvement of bed day utilization using Bradford Hills criteria [22]. The effect size was 1.14 bed days saved per person, yet only 43.6% of the intention-to-treat group (471/1080) participated in the intervention suggesting a strong effect of the intervention. Positive effects using the Patient Journey Record System component of the MonashWatch model were demonstrated in 3 Irish cohorts [11] and another Victorian location [23]. MonashWatch has been judged by the Victorian DHHS to have improved efficiencies, with better care and external evaluation supporting a significant improvement in bed days of around 0.8 days per person [20]. Other evidence of a causal relationship is that a more significant reduction in bed days occurred immediately after entry into MonashWatch by the intention-to-treat group and persisted throughout the evaluation.

Why Did the Intervention Appear to Work?

A plausible mechanism for the intervention working is addressing resilience and frailty through anticipatory care and coaching enabling stronger health networks and connections with vulnerable people. The results support a continuous adaptation to complex unstable health journey model for individuals [24]. Many hospital admissions, even very short stays in the emergency department (which are included in this study as an admission if there is an overnight stay) may be related to a wide range of indirect influences [25]. Underlying issues could include personality types, mood, anxiety or demoralization, drug and alcohol or medication management problems, service access issues and convenience, as well as the likelihood of benefit from hospital-level care [26,27]. A sense of discrimination has been described in emergency department frequent users [27]. Many of these factors (but not all) can be addressed through closer monitoring in a prehospital phase [28]. It is worth noting that a high net-promoter rate (satisfaction scores) of 95% was consistently reported by patients participating in MonashWatch at 6-month intervals and on exit. The accompanying narrative commentary was very positive about feeling “accepted” and “supported.” This commentary will be reported elsewhere and makes frequent mention of improved health experience and a sense that the hospital “cares.” This also suggests we are heading in the right direction.

The capitated structure provided very adaptable funding as needed for many issues such as transport, outpatient attendances, and home factors which underpin admissions. It provided coaches with the flexibility to go outside of health siloes working with general practitioners, hospital, alcohol, and mental health services.

Finally, this DHHS approach that is outcome and data-driven with continuous performance and costing review for teams and local initiatives keeps services from falling into complacency.

Limitations

There is a range of limitations. The intention-to-treat group included 44.9% (485/1080) who declined. This arduous process may have diluted the uptake rate and thus had an impact on bed days but reflects real-world clinical service evaluations. It would be worth other methods of recruitment in the future to see if there would be increased recruitment with a more significant impact.

Pragmatic clinical evaluations in live health systems outside of research study are challenging, particularly in MonashWatch due to long-term unpredictable, complex dynamics in unstable patient journeys [14]. There are practical difficulties in finding true controls beyond the HLCC scoring algorithm. Systematic contemporaneous and ecological selection within a local geographical zone, as in this case, was the option judged to be most reliable. Retrospective propensity scoring based on HLCC, further stratified by diagnoses, age, and socioeconomic (postal code) status with randomly selected multiple controls in different hospitals, was employed by the external evaluation [20]. However, the findings of the MonashWatch approach were mirrored with a significant 0.8 bed day saving per person in the external evaluation [20]. Resource utilization outcome measures and the actual rate of savings will be the subject of further evaluation.

This summative evaluation was conducted in a living health system as the first phase of a government funding initiative to move from activity to value-based funding when Monash Health services were under significant funding constraints. Success was achieved in the real world without going through the traditional research route with a trial before rolling out an implementation. The positive feature of this approach is that (to date) it has not gone the way of many beautifully designed and executed pilots that never achieved implementation. In the first phase, the successful delivery of care was within existing funding. The MonashWatch-type model deployment in other health services in the second and third phases has the opportunity to improve on trial methodology. The addition of more research resources, given the current successful proof of concept, would enable the conduct of a more sophisticated randomized propensity-matched trial.

Ongoing study of the data is needed to identify who benefits from which components and how the intervention can be improved for different groups. There is a need to continue to shift current care pathways and health systems to adapt care to the needs of vulnerable MonashWatch-type patients. A whole of systems transformation is needed to respond to the dynamics of unstable health journeys, beyond the current single disease or condition siloed care. Outcome-based funding has the potential to make major inroads into fragmented care. Macrolevel health service funding changes in the Australian Health System such as in northern Spain [29] would be a great advantage to improve integrated care but is unlikely due to political barriers.

Conclusion

The MonashWatch telehealth and coaching intervention using the HLCC innovative funding model was effective in a local catchment area of a hospital in a highly disadvantaged community and achieved its health service funding model objectives. It requires ongoing and broader implementation and evaluation. In the future, evaluation of additional teams is needed to confirm these findings in different populations and settings. Two additional teams are now in place. Ultimately, the progressive scaling up to a multisite intervention will require ongoing tailoring and evaluation with feedback for improvement.

Acknowledgments

This paper acknowledges the innovative funding model (HLCC) developed by the Victorian Department of Health and Human Services, and the very generous open spirit of the HLCC team. It acknowledges the stellar work of the MonashWatch clinical team—the telecare guides and the coaches who have made the model work to date in the highly charged domain of living health services. It acknowledges the work of Kevin Smith and John-Paul Smith who have implemented and supported the Patient Journey Record System app and analytics. Kevin Smith has provided ongoing comments, editing and proofreading support during the writing of this paper. John Kellett is acknowledged as a particularly thorough reviewer. DHHS continually tracks the data and has an independent summative evaluation being conducted by the Commonwealth Scientific and Industrial Research Organization. Smart Health Solutions have developed and implemented the web-based system for supporting coaching of the patients identified by Patient Journey Record System in their PatientWatch system.

Authors' Contributions

KS and CM both analyzed the data separately and independently from each other.

Conflicts of Interest

CM was the clinician lead in the development of the Patient Journey Record System software.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 974 KB - jmir_v22i12e18046_app1.pdf]

References

1. Australian Institute of Health and Welfare (AIHW). Selected potentially preventable hospitalisations, P1 18. National Health Care Agreement. Metadata Online Registry. Canberra, Australia: Australian Government; 2018 Jan 30. URL: https://meteor.aihw.gov.au/content/index.phtml/itemId/658499 [accessed 2020-08-14]
2. Ferrier D, Diver F, Corin S, McNair P, Cheng C. HealthLinks: Incentivising better value chronic care in Victoria. Int J Integr Care 2017 Jul 11;17(3):129 [FREE Full text] [doi: 10.5334/ijic.3241]
3. Segal L, Dunt D, Day SE. Introducing coordinated care (2): evaluation of design features and implementation processes implications for a preferred health system reform model. Health Policy 2004 Aug;69(2):215-228. [doi: 10.1016/j.healthpol.2004.02.001] [Medline: 15212868]
4. Gardner K, Yen L, Banfield M, Gillespie J, McRae I, Wells R. From coordinated care trials to medicare locals: what difference does changing the policy driver from efficiency to quality make for coordinating care? Int J Qual Health Care 2013 Feb;25(1):50-57. [doi: 10.1093/intqhc/mzs069] [Medline: 23175532]
5. How much does Australia spend on health care? Australia’s health 2018. Australia’s health series no. 16. AUS 221. Canberra, Australia: Australian Institute of Health and Welfare; 2018 Jun 20. URL: https://www.aihw.gov.au/getmedia/941d2d8b-68e0-4883-a0c0-138d43dba1b0/aihw-aus-221-chapter-2-2.pdf.aspx [accessed 2020-10-13]
6. Victorian Initiatives, Hospital Admission Risk Program in Integrated Care,. Shifting the Dial 5 Year Productivity Review. Canberra: Australian Government Productivity Commission; 2017 Oct 17. URL: https://www.pc.gov.au/inquiries/completed/productivity-review/report [accessed 2020-10-13]
7. Ferrier D, Campbell D, Hamilton C, McNair P, Rocca R, Sewards T, et al. Co-designing a new approach to delivering integrated services to chronically ill patients within existing funding constraints – Victoria’s HealthLinks trial. Int J Integr Care 2019 Aug 08;19(4):322 [FREE Full text] [doi: 10.5334/ijic.s3322]
8. Giovannetti ER, Dy S, Leff B, Weston C, Adams K, Valuck TB, et al. Performance measurement for people with multiple chronic conditions: conceptual model. Am J Manag Care 2013 Oct 01;19(10):e359-e366 [FREE Full text] [Medline: 24304183]

9. Bayliss EA, Ellis JL, Steiner JF. Seniors' self-reported multimorbidity captured biopsychosocial factors not incorporated into two other data-based morbidity measures. J Clin Epidemiol 2009 May;62(5):550-7.e1 [FREE Full text] [doi: 10.1016/j.jclinepi.2008.05.002] [Medline: 18757178]
10. Wister AV, Coatta KL, Schuurman N, Lear SA, Rosin M, MacKey D. A lifecourse model of multimorbidity resilience: theoretical and research developments. Int J Aging Hum Dev 2016 Apr;82(4):290-313. [doi: 10.1177/0091415016641686] [Medline: 27076489]
11. Martin CM, Vogel C, Grady D, Zarabzadeh A, Hederman L, Kellett J, et al. Implementation of complex adaptive chronic care: the Patient Journey Record system (PaJR). J Eval Clin Pract 2012 Dec;18(6):1226-1234. [doi: 10.1111/j.1365-2753.2012.01880.x] [Medline: 22816797]
12. Martin CM. Self-rated health: patterns in the journeys of patients with multi-morbidity and frailty. J Eval Clin Pract 2014 Dec;20(6):1010-1016. [doi: 10.1111/jep.12133] [Medline: 24828245]
13. Ackoff RL. New tools, new thinking. Interview by Joe Flower. Healthc Forum J 1992;35(2):62-67. [Medline: 10117132]
14. Martin CM, Sturmberg JP, Stockman K, Hinkley N, Campbell D. Anticipatory care in potentially preventable hospitalizations: making data sense of complex health journeys. Front Public Health 2018;6:376 [FREE Full text] [doi: 10.3389/fpubh.2018.00376] [Medline: 30746358]
15. Ćirković S. Buurtzog - revolutionising home care in the Netherlands. Centre for Public Impact, A BCG Foundation. 2018 Nov 15. URL: https://www.centreforpublicimpact.org/case-study/buurtzorg-revolutionising-home-care-netherlands/ [accessed 2020-10-14]
16. Victorian Admitted Episodes Dataset (VAED). Better Safer Care Victoria. Melbourne, Victoria: Department of Health; 2012. URL: https://www.bettersafercare.vic.gov.au/sites/default/files/2018-09/Victorian%20Admitted%20Episodes%20Dataset%20_VAED_%20Accessible%20%26%20Restricted%20Data%20Fields.pdf [accessed 2020-10-14]
17. Analysis of Covariance (ANCOVA). XLSTAT. Paris, France: Addinsoft; 2020 Apr 01. URL: https://www.xlstat.com/en/solutions/features/ancova-analysis-of-covariance [accessed 2020-10-14]
18. Patient Experience. Healthecare Australia. Sydney, Australia; 2018. URL: https://healthecare.com.au/quality/patient-experience [accessed 2020-05-05]
19. Adjusting for Covariates in Randomised Clinical Trials for Drugs and Biologics with Continuous Outcomes Guidance for Industry. U.S. Department of Health and Human Services Food and Drug Administration. 2019 Feb. URL: https://www.fda.gov/media/123801/download [accessed 2019-11-13]
20. HealthLinks Clinical Collaborative Group (HLCC). Evaluation Results from 2016-17 and 2017-18. In: Victorian Department of Health and Human Services. 2019 Presented at: Interim Evaluation Meeting; August 21 2019; Melbourne p. 8.
21. Christensen CM, Bohmer R, Kenagy J. Will disruptive innovations cure health care? Harv Bus Rev 2000;78(5):102-12, 199. [Medline: 11143147]
22. Hill AB. The Environment and Disease: Association or Causation? Proceedings of the Royal Society of Medicine 2016 Sep;58(5):295-300 [FREE Full text] [doi: 10.1177/003591576505800503]
23. Murphy A, Martin C, Jennifer C, Cooper C, Kumnick A, Hallissey S. The Healthy@Home Pilot Service Final Report. In: Evaluation of Healthy@Home East Grampians Health Services. Victoria, Australia: Centre for eResearch and Digital Innovation, Federation University; 2020.
24. Martin C, Sturmberg J. Complex adaptive chronic care. J Eval Clin Pract 2009 Jun;15(3):571-577 [FREE Full text] [doi: 10.1111/j.1365-2753.2008.01022.x] [Medline: 19522913]
25. Šteinmiller J, Routasalo P, Suominen T. Older people in the emergency department: a literature review. Int J Older People Nurs 2015 Dec;10(4):284-305 [FREE Full text] [doi: 10.1111/opn.12090] [Medline: 26183883]
26. Baggio S, Iglesias K, Hugli O, Burnand B, Ruggeri O, Wasserfallen J, et al. Associations between perceived discrimination and health status among frequent emergency department users. Eur J Emerg Med 2017 Apr;24(2):136-141. [doi: 10.1097/MEJ.0000000000000311] [Medline: 26267073]
27. Pinchbeck E. Convenient primary care and emergency hospital utilisation. J Health Econ 2019 Dec;68:102242 [FREE Full text] [doi: 10.1016/j.jhealeco.2019.102242] [Medline: 31605834]
28. Martin C, Hinkley N, Stockman K, Campbell D. Potentially preventable hospitalizations-The 'pre-hospital syndrome': Retrospective observations from the MonashWatch self-reported health journey study in Victoria, Australia. J Eval Clin Pract 2020 Aug 28:1-8. [doi: 10.1111/jep.13460] [Medline: 32857482]
29. Bengoa R. Transforming health care: an approach to system-wide implementation. Int J Integr Care 2013;13:e039 [FREE Full text] [doi: 10.5334/ijic.1206] [Medline: 24179458]

Abbreviations

AIHW: Australian Institute of Health and Welfare
ANCOVA: analysis of covariance
DHHS: Department of Health and Human Services
HLCC: HealthLinks: Chronic Care

ICD: International Statistical Classification of Diseases

Edited by G Eysenbach; submitted 30.01.20; peer-reviewed by J Kellett, B Cotner; comments to author 27.03.20; revised version received 27.05.20; accepted 11.08.20; published 01.12.20.

Please cite as:
Martin C, Hinkley N, Stockman K, Campbell D
Capitated Telehealth Coaching Hospital Readmission Service in Australia: Pragmatic Controlled Evaluation
J Med Internet Res 2020;22(12):e18046
URL: https://www.jmir.org/2020/12/e18046
doi:10.2196/18046
PMID:33258781

Original Paper

People's Experiences and Satisfaction With Telehealth During the COVID-19 Pandemic in Australia: Cross-Sectional Survey Study

Jennifer MJ Isautier[1], MSc; Tessa Copp[1], PhD; Julie Ayre[1], PhD; Erin Cvejic[1], PhD; Gideon Meyerowitz-Katz[2,3], MPH; Carys Batcup[1], MSc; Carissa Bonner[1], PhD; Rachael Dodd[1], PhD; Brooke Nickel[1], PhD; Kristen Pickles[1], PhD; Samuel Cornell[1], MSc; Thomas Dakin[1], MPH; Kirsten J McCaffery[1], PhD

[1]Faculty of Medicine and Health, Sydney Health Literacy Lab, School of Public Health, The University of Sydney, Sydney, Australia
[2]Population Wellbeing and Environment Research Lab, School of Health and Society, University of Wollongong, Wollongong, Australia
[3]Western Sydney Diabetes, Western Sydney Local Health District, Sydney, Australia

Corresponding Author:
Jennifer MJ Isautier, MSc
Faculty of Medicine and Health
Sydney Health Literacy Lab, School of Public Health
The University of Sydney
Edward Ford Building
A27 Fisher Rd
Sydney, 2006
Australia
Phone: 61 02 9114 2199
Email: jennifer.isautier@sydney.edu.au

Abstract

Background: In response to the COVID-19 pandemic, telehealth has rapidly been adopted to deliver health care services around the world. To date, studies have not compared people's experiences with telehealth services during the pandemic in Australia to their experiences with traditional in-person visits.

Objective: This study aimed to compare participants' perceptions of telehealth consults to their perceptions of traditional in-person visits and investigate whether people believe that telehealth services would be useful after the pandemic.

Methods: A national, cross-sectional, community survey was conducted between June 5 and June 12, 2020 in Australia. In total, 1369 participants who were aged ≥18 years and lived in Australia were recruited via targeted advertisements on social media (ie, Facebook and Instagram). Participants responded to survey questions about their telehealth experience, which included a free-text response option. A generalized linear model was used to estimate the adjusted relative risks of having a poorer telehealth experience than a traditional in-person visit experience. Content analysis was performed to determine the reasons why telehealth experiences were worse than traditional in-person visit experiences.

Results: Of the 596 telehealth users, the majority of respondents (n=369, 61.9%) stated that their telehealth experience was "just as good as" or "better than" their traditional in-person medical appointment experience. On average, respondents perceived that telehealth would be moderately useful to very useful for medical appointments after the COVID-19 pandemic ends (mean 3.67, SD 1.1). Being male (P=.007), having a history of both depression and anxiety (P=.016), and lower patient activation scores (ie, individuals' willingness to take on the role of managing their health/health care) (P=.036) were significantly associated with a poor telehealth experience. In total, 6 overarching themes were identified from free-text responses for why participants' telehealth experiences were poorer than their traditional in-person medical appointment experiences, as follows: communication is not as effective, limitations with technology, issues with obtaining prescriptions and pathology results, reduced confidence in their doctor, additional burden for complex care, and inability to be physically examined.

Conclusions: Based on our sample's responses, telehealth appointment experiences were comparable to traditional in-person medical appointment experiences. Telehealth may be worthwhile as a mode of health care delivery while the pandemic continues, and it may continue to be worthwhile after the pandemic.

(J Med Internet Res 2020;22(12):e24531) doi:10.2196/24531

KEYWORDS

COVID-19; patient experience; telehealth; experience; satisfaction; telemedicine; Australia; usability; cross-sectional; survey

Introduction

The COVID-19 outbreak was officially declared a pandemic by the World Health Organization on March 11, 2020. To help minimize the spread of COVID-19, health care systems have rapidly adopted alternative models for health care delivery, including telehealth services [1]. This type of health care delivery minimizes the spread of the virus by providing health care services without the need for close contact, thereby reducing the risk of exposure to COVID-19 for both patients and clinicians.

In response to the COVID-19 pandemic, the Australian Government introduced a temporary telehealth scheme on March 30, 2020 to enable subsidized access to health care services that are provided via telephone or videoconferencing [2]. Prior to the pandemic, telehealth consultations were restricted to rural and remote communities. This new scheme has allowed all medical appointments with a variety of health professionals to be conducted via telehealth, regardless of rurality. As a result of this scheme, telehealth consults have accounted for 36% of all services provided in April 2020, whereas telehealth consults conducted before the pandemic only accounted for 1.3% [3,4]. At the end of April 2020, a nationally representative survey of 1022 people conducted by the Australian Bureau of Statistics reported that 1 in 6 people (17%) used a telehealth service, women were almost twice as likely as men to use telehealth services (22% vs 12%), and persons with a chronic or mental health condition were twice as likely to have used a telehealth service compared to those without such conditions (25% vs 13%). However, 1 in 10 people (10%) reported to have a general practitioner or health professional appointment cancelled or postponed in the last 4 weeks because of the COVID-19 pandemic [5].

Cancelling or postponing appointments is concerning because reduced health care during pandemics has been associated with poor health outcomes, as observed during the Ebola virus outbreak and Severe acute respiratory syndrome epidemic [6,7]. The increased uptake of telehealth services and increased number of people cancelling or postponing medical appointments warrants further investigation to better understand people's experiences and satisfaction with accessing telehealth services during the COVID-19 pandemic. This is particularly necessary, given the long-term outlook of the COVID-19 pandemic; although several health services have returned to normal, continuing outbreaks may deter patients from accessing in-person care for some time [8].

Despite the growth of telehealth, no studies have compared people's experiences with telehealth services during the COVID-19 pandemic in Australia to people's experiences with traditional in-person visits. We investigated a sample of Australians and their experiences with telehealth during the COVID-19 pandemic. Our aims were to compare participants' perceptions of telehealth consults to their perceptions of traditional in-person visits and investigate whether people believe that telehealth services would be useful after the pandemic. Furthermore, we investigated the sociodemographic and health-related factors associated with negative telehealth experiences.

Methods

Recruitment

The data used in this study are from a prospective, longitudinal, national survey that launched in April 2020 and explored variations in people's understanding of, attitude toward, and uptake of COVID-19 health advice during the 2020 pandemic [9]. Herein, we report on data from a survey wave conducted over a 1-week period (ie, June 5 to June 12, 2020) in Australia. Data were obtained using the Qualtrics online platform. Participants who were aged ≥18 years, could read and understand English, and resided in Australia were recruited via paid targeted advertisements on social media (ie, Facebook and Instagram). More details on recruitment are provided in the McCaffery et al study [9]. Participants were given the opportunity to enter a prize draw for the chance to win 1 of 10 Aus $20 (US $14.62) gift cards upon completion of the survey. This study was approved by The University of Sydney Human Research Ethics Committee (2020/212).

Measures

Sociodemographic variables, including age, gender, and educational status, were collected, along with data on self-reported chronic diseases and overall health. We assessed health literacy using the Newest Vital Sign [10] and digital health literacy using the eHealth Literacy Scale [11]. The Consumer Health Activation Index [12] was used to determine patient activation (ie, individuals' willingness to take on the role of managing their health and health care). The remoteness and socioeconomic status of participants' places of residence were derived from participants' postcodes [13]. Participants were asked to indicate whether they had used telehealth services. If so, they were then asked how telehealth services compared to traditional in-person visits, whether they experienced any barriers to using telehealth services, and whether they cancelled or postponed an appointment with a health professional (Textbox 1).

Textbox 1. Survey items and scoring scale on telehealth.

Telehealth usage

- Since the COVID-19 restrictions started, have you had a telemedicine/telehealth appointment (appointment with your health provider by video or phone instead of an in-person appointment)? (Response options: Yes/No)
- How many telehealth appointments have you had? (Response: Numerical [free-text])
- Was/were your telemedicine/telehealth visit(s) done by: (Response options: Telephone/Videoconference/Both)

Comparison between telehealth and traditional in-person visits

- How did your telemedicine/telehealth visit compare to a traditional in-person medical visit? (Response options: Better than a traditional visit/Just as good as a traditional visit/Worse than a traditional visit/Not sure)
- If, telemedicine/telehealth was worse, please tell us why. (Response: Free text)

Interest in telehealth after COVID-19

- How useful do you think it will be to have medical appointments with telemedicine/telehealth after the COVID-19 emergency is over? (Response scale: 1-5, indicating not at all to extremely)

Cancellation of in-person appointments

- Have you cancelled or postponed an appointment with a health professional in the last 4 weeks because of COVID-19? (Response options: Yes/No)
- Why? (Response options: Concerned about the cost/I am isolating due to COVID-19 symptoms or risk/I was worried about travelling on public transport because of the COVID-19 risk/I did not want to go to a health or hospital clinic because of concerns about catching COVID-19 there/Too busy Other [please tell us])
- Did you feel you needed to see a health professional in person in the last 4 weeks but chose not to go? (Response options: Yes/No)
- Why? (Response options: Concerned about the cost/I am isolating due to COVID-19 symptoms or risk/I was worried about travelling on public transport because of the COVID-19 risk/I did not want to go to a health or hospital clinic because of concerns about catching COVID-19 there/Too busy/Other [please tell us])

Barriers to telehealth

- Have you needed to access a telehealth service in the last 4 weeks but could not? (Response options: Yes/No)
- What was the main reason that you could not access a telehealth service in the last 4 weeks? (Response options: Telehealth not available from general practitioner or other health professional/Do not have internet/I am not able to use the internet/Dislike or fear of the service/Appointment not available when required/Other [please detail])

Statistical Analysis

Quantitative data were analyzed using Stata/IC v16 (StataCorp LLC). Descriptive statistics were analyzed to obtain sample characteristics and summarize participants' telehealth experiences since COVID-19 restrictions commenced. A generalized linear model using a modified Poisson approach (ie, log link function with robust standard errors) was used to estimate adjusted relative risks with 95% confidence intervals for having a poorer telehealth experience than a traditional in-person medical visit experience based on various sociodemographic and health-related factors. A 2-tailed independent samples *t* test was used to compare the perceived usefulness of telehealth medical appointments once the COVID-19 emergency ends between participants who rated their telehealth experience as worse than their in-person medical visit experience and those who rated their telehealth experience as the same as or better than their in-person medical visit experience. The statistical significance for these exploratory analyses was set at *P*<.05 (2-tailed).

Qualitative data were analyzed using content analysis [14], which combines both qualitative and quantitative methods and allows for both the frequency of categories and the content to be reported. JI and TC familiarized themselves with the content and generated a list of recurring themes; these were discussed with and checked by an additional researcher (JA). JI and TC then applied the final coding framework to all the data. The level of agreement was tested using the Cohen kappa, which indicated substantial agreement (κ=0.76) [15]. Discrepancies were discussed until a consensus was obtained. Descriptive statistics were provided to summarize the frequency of each code.

Results

Of the 1369 respondents who completed the June survey, 596 (43.5%) reported using telehealth services since the start of the pandemic. Respondents who used telehealth services were slightly older; more likely to be female; had higher levels of education; had a greater prevalence of chronic health conditions, including a history of mental health conditions; and had poorer self-reported general health compared to those who did not use telehealth services. Sample characteristics are summarized in Table 1.

Table 1. Descriptive characteristics of our sample sorted by participants' use of telehealth services during the COVID-19 lockdown period.

Variable	Accessed telehealth services (n=596)	Did not access telehealth services (n=773)	Overall (N=1369)
Age in years, mean (SD)	46.2 (16.1)	43.6 (17.0)	44.7 (16.7)
Age group (years), n (%)			
18-25	76 (12.8)	156 (20.2)	232 (16.9)
26-40	166 (27.9)	206 (26.6)	372 (27.2)
41-55	152 (25.5)	192 (24.8)	344 (25.1)
56-90	202 (33.9)	219 (28.3)	421 (30.8)
Gender, n (%)			
Male	146 (24.5)	287 (37.1)	433 (31.6)
Female	433 (72.7)	478 (61.8)	911 (66.5)
Other/prefer not to say	17 (2.9)	8 (1)	25 (1.8)
Highest level of education completed, n (%)			
High school or less	68 (11.4)	130 (16.8)	198 (14.5)
Certificate I-IV	67 (11.2)	73 (9.4)	140 (10.2)
University education	461 (77.3)	570 (73.7)	1031 (75.3)
Number of chronic health conditions[a], n (%)			
0	239 (40.1)	436 (56.4)	675 (49.3)
1	188 (31.5)	220 (28.5)	408 (29.8)
≥2	169 (28.4)	117 (15.1)	286 (20.9)
Mental health history, n (%)			
Depression	278 (46.6)	193 (25.0)	471 (34.4)
Anxiety	302 (50.7)	232 (30)	534 (39)
Self-reported general health, n (%)			
Poor	37 (6.2)	9 (1.2)	46 (3.4)
Fair	111 (18.6)	76 (9.8)	187 (13.7)
Good	226 (37.9)	237 (30.7)	463 (33.8)
Very Good	172 (28.9)	321 (41.5)	493 (36)
Excellent	50 (8.4)	130 (16.8)	180 (13.1)
Socioeconomic status, mean IRSAD[b] quintile (SD)	3.7 (1.4)	3.7 (1.4)	3.7 (1.4)
Remoteness, n (%)			
Major cities	438 (73.5)	589 (76.2)	1027 (75)
Other	158 (26.5)	184 (23.8)	342 (25)
Adequate health literacy[c], n (%)	505 (90.3)	665 (91.7)	1170 (91.1)
eHealth literacy[d], mean (SD)	4.2 (0.7)	4.1 (0.7)	4.2 (0.7)
Patient activation[e], mean (SD)	74.7 (13.2)	75.0 (13.4)	74.9 (13.3)
Cancelled/postponed an appointment[f], n (%)	147 (24.7)	125 (16.2)	272 (19.9)
Chose not to see a health professional[g], n (%)	115 (19.3)	104 (13.5)	219 (16)
Could not access telehealth services[h], n (%)	12 (2)	7 (0.9)	19 (1.4)

[a] Chronic health conditions included respiratory disease, asthma, chronic obstructive pulmonary disease, hypertension, cancer, heart disease, stroke, and diabetes.

[b] IRSAD: Index of Relative Socio-Economic Advantage and Disadvantage. In the IRSAD quintile [13], a score of 1 represents most disadvantaged and

a score of 5 represents most advantaged.

[c] Health literacy was assessed using the Newest Vital Sign [10]. Data were missing for 85 (6.2%) participants percent due to technical errors with the Qualtrics online platform.

[d] eHealth [11] literacy was measured on a 5-point Likert scale. A higher score reflects a higher level of eHealth literacy.

[e] Results are based on the Consumer Health Activation Index [12]. A score of 0-79 indicates low activation, 80-94 indicates moderate activation, and 95-100 indicates high activation.

[f] Respondents who cancelled/postponed an appointment in the last 4 weeks because of COVID-19.

[g] Respondents who felt the need to see a health professional in the last 4 weeks, but chose not to.

[h] Respondents who needed access to a telehealth service in the last 4 weeks, but could not.

Cancellation of In-Person Appointments

Of the 1369 total respondents, 272 (19.9%) cancelled or postponed an in-person appointment with a health professional. The reasons for cancelling appointments were as following: concerns about catching COVID-19 at a clinic or hospital (n=85, 31.3%), isolating due to COVID-19 symptoms or risks (n=31, 11.4%), concerns about travelling via public transport (n=21, 7.7%), feeling too busy (n=20, 7.4%), cost (n=9, 3.3%), and other reasons (n=106, 39%). Less common reasons for cancelling or postponing an in-person appointment included the following: border closures, postponed elective surgery, and the appointment seemed nonessential. Furthermore, 219 (16%) respondents felt that they needed to see a health professional in-person in the last 4 weeks, but chose not to go due to the following reasons: concerns about catching COVID-19 at a clinic or hospital (n=72, 32.9%), feeling too busy (n=37, 16.9%), isolating due to COVID-19 symptoms or risks (n=18, 8.2%), concerns about travelling via public transport (n=13, 5.9%), other reasons (n=67, 30.6%). Less common reasons listed for choosing not to see a health professional included the following: only telehealth services were available, limited in-person appointment availability, and felt that seeing a health professional was too complicated.

Telehealth Experiences

The characteristics of telehealth users' experiences are shown in Table 2. Of the 596 respondents who used telehealth services, over half (n=326, 54.7%) reported having more than 1 telehealth appointment, of which most were conducted by telephone (n=427, 71.6%). The majority of respondents (n=369, 61.9%) stated that their telehealth experience was "just as good as" or "better than" their traditional in-person medical visit experience. On average, respondents perceived telehealth as moderately useful to very useful for medical appointments after the COVID-19 pandemic ends (mean 3.67, SD 1.1). Individuals who responded that their telehealth experience was worse than their traditional in-person medical visit experience (n=205, 34.4%) also rated the usefulness of telehealth after the COVID-19 emergency ends significantly lower than those whose telehealth experience was "just as good as" or "better than" their in-person visit experience (mean 2.86 vs mean 4.17; difference: mean 1.31; 95% CI 1.14-1.47; t_{572}=15.62; P<.001).

Table 2. Characteristics of telehealth users' experience (n=596).

Variable	Summary value, n (%)
Number of telehealth appointments	
1	270 (45.3)
2	157 (26.3)
≥3	169 (28.4)
Mode of telehealth delivery	
Telephone	427 (71.6)
Videoconference	84 (14.1)
Both	85 (14.3)
Telehealth visit compared to traditional in-person medical visit	
Better	49 (8.2)
Just as good	320 (53.7)
Worse	205 (34.4)
Unsure	22 (3.7)

Factors Associated With a Poor Telehealth Experience

The results of the multivariable analysis that explored factors associated with a poorer telehealth experience than an in-person appointment experience are displayed in Table 3. Being male (*P*=.007), having a history of both depression and anxiety (*P*=.04), and having a low patient activation score (*P*=.04) were associated with a poorer telehealth experience, after controlling for all other variables in the model.

Table 3. Multivariable[a] analysis of factors associated with a poorer[b] telehealth experience than an in-person appointment experience (n=574). Adjusted relative risks of <1 indicate a reduced risk of reporting a poorer telehealth experience relative to the reference group.

Variable	Adjusted relative risk (95% CI)	*P* value
Age (years)[c]		.27
18-25	0.98 (0.66-1.47)	.94
26-40	1.09 (0.80-1.49)	.57
56-90	1.32 (0.97-1.80)	.08
Gender[d]		.01
Female	0.73 (0.58-0.92)	.007
Other/prefer not to say	0.52 (0.24-1.14)	.11
Highest level of education completed[e]		.96
Certificate I-IV	1.06 (0.67-1.67)	.80
University education	1.01 (0.69-1.48)	.96
Number of chronic health conditions[f]		.26
1	0.88 (0.68-1.15)	.35
≥2	0.78 (0.58-1.05)	.11
Mental health history[g]		.054
Either depression or anxiety	1.27 (0.92-1.75)	.14
Both depression and anxiety	1.42 (1.07-1.89)	.016
Socioeconomic status (per IRSAD[h] quintile)	1.06 (0.97-1.16)	.20
eHealth literacy (per unit)	1.02 (0.84-1.23)	.84
Patient activation (per 10-unit increase)	0.91 (0.82-0.99)	.036
Telehealth delivery mode[i]		.23
Videoconference	1.28 (0.96-1.70)	.09
Both telephone and videoconference	1.09 (0.79-1.51)	.60

[a] The analysis also controlled for the number of telehealth visits since lockdown.

[b] A poorer outcome was defined as respondents rating their telehealth experience as worse (compared to “just as good as” or “better than”) than their traditional in-person medical visit experience. Individuals who responded with “unsure” (22/596, 3.7%) were excluded from the analysis.

[c] Respondents aged 41-55 years were used as a reference.

[d] Male respondents were used as a reference.

[e] Respondents who completed a high school education or less were used as a reference.

[f] Respondents who did not have a chronic health condition were used as a reference.

[g] Respondents who did not have a history of mental health conditions were used as a reference.

[h] IRSAD: Index of Relative Socio-Economic Advantage and Disadvantage.

[i] Respondents who had telehealth visits via telephone were used as a reference.

Reasons Provided for Why Telehealth Experiences Were Worse Than Traditional In-Person Visit Experiences

In total, the following 6 overarching themes regarding telehealth services (Table 4) emerged from the free-text data: (1) communication is not as effective as face-to-face visits, (2) limitations with technology, (3) issues with obtaining prescriptions and pathology results, (4) reduced confidence in the doctor, (5) additional burden for complex care, and (6) inability to be physically examined. Of the 221 respondents who provided a free-text response for why their telehealth experiences were worse than their traditional in-person visit experiences, the majority received telehealth services via telephone (n=149, 67.4%), whereas comparatively fewer respondents received telehealth services via videoconference (n=37, 16.7%) or a combination of both (n=35, 15.8%). However, no substantial differences in the overarching themes were observed between these groups. Overall, the most recurrent theme was that communication was not as effective as traditional in-person visits due to the lack of visual cues, eye contact, and body language.

Table 4. Reasons provided by 221 respondents for telehealth visits being worse than traditional in-person medical visits, along with the frequency of overarching themes and subthemes with example quotes[a].

Code description	Example	Value, n (%)
Communication is not as effective as face-to-face visits		
Lacks visual cues, eye contact, body language, and visual feedback; face-to-face visits are preferred	"The subtle facial expressions eye contact and body language are not the same"	54 (24.4)
Less personal, less natural/comfortable, more awkward	"Difficult to establish rapport"	46 (20.8)
Less effective; communication is not as good, less helpful, and harder/more difficult	"Communication on the phone is less effective"	45 (20.4)
Face-to-face interaction is needed for mental health appointments	"I feel a big part of effective mental health care involves face-to-face conversation"	19 (8.6)
More anxiety provoking for some	"Phone call and videos make me extremely anxious"	5 (2.3)
Less privacy	"Lack of privacy"	3 (1.4)
The inability to be physically examined		
Physical exam is not possible	"Could not have a physical exam done"	60 (27.1)
Tests could not be performed	"Blood pressure not taken"	17 (7.7)
Limitations with technology		
Technology issues, including poor connection, bad reception, poor audio quality, and Zoom calls dropping out	"Due to audio quality I was not able to get names of chemotherapy drugs correctly - so when I tried to look up info later I couldn't until I was able to get info from Breast care nurse so this added to days of anxiety due to lack of info over weekend and when that staff member on leave."	20 (9)
Poor quality connection led to poor quality conversation	"Harder to communicate due to tech difficulties, lag issues"	6 (2.7)
Issues with obtaining prescriptions and pathology results		
Harder to obtain prescriptions	"I had to wait for scripts to be emailed to the pharmacy, then one was missing, which I could have seen at the time had I received them in person."	10 (4.5)
Increased wait time/delayed access	"If you need a script or referral, you have to make a separate trip to go get the paper"	7 (3.2)
Unable to access pathology results	"Getting blood tests has become more difficult."	2 (0.9)
Reduced confidence in doctor/health professional		
Not as comprehensive or thorough	"Not as comprehensive and thorough"	25 (11.3)
Time pressure	"Felt rushed"	18 (8.1)
Lack of confidence in assessment/ diagnosis	"Less trust that the diagnosis is accurate"	11 (5)
Additional burden for complex care		
Face-to-face visit required due to complex issues	"I had to go in for a face-face consult because the medical issues could not be diagnosed"	15 (6.8)
Delays due to complex issues	"More complex issues have been delayed until we can do face-to-face"	5 (2.3)
Added burden of having 2 consults	"In both instances after having the Telehealth calls, I had to go in for a face to face consults because the issues could not be diagnosed over the phone"	4 (1.8)

[a] Full text could have more than 1 theme applied.

Barriers to Telehealth

Of the 1369 total respondents, 19 (1.4%) reported that they were not able to access a telehealth service. The barriers reported by these 19 respondents were as follows: telehealth services were not available from their general practitioner or health professional (n=4, 21.1%), they did not have internet (n=2, 10.5%), the appointment was not available when required (n=8, 42.1%), and using telehealth services felt too complicated (n=5, 26.3%).

Discussion

Principal Results

Our findings showed that more than half of the respondents (n=369, 61.9%) stated that their telehealth experience was “just as good as” or “better than” their traditional in-person medical care experience. This is encouraging, considering that the community transmission of COVID-19 across Australia may continue to persist for some time. On average, respondents perceived that telehealth would be moderately useful to very useful for medical appointments after the COVID-19 pandemic. This suggests that telehealth may be a viable long-term option for health care delivery. Our findings are consistent with another survey, which reported that 85% of older Australians found their telehealth experiences to be similar or better than their experiences with face-to-face consults [16]. Furthermore, we found that telehealth delivery modality (ie, telephone and video) was not associated with having a poorer telehealth experience than an in-person appointment experience (*P*=.23). This is consistent with other studies that showed telephone and videoconferencing were comparable in terms of patient satisfaction [17]. It is perhaps unsurprising that respondents who rated their telehealth experiences as worse than their traditional in-person visit experiences were less likely to perceive that telehealth would be useful after the COVID-19 pandemic. Such respondents were more likely to be male, have lower patient activation scores (ie, individuals’ willingness to take on the role of managing their health and health care), and have a history of both depression and anxiety (Table 3). This last observation is also supported by the content analysis, which highlighted that several participants preferred in-person to telehealth visits for mental health appointments.

Limitations

While the study sample was large and diverse, it was not statistically representative of the Australian population. Our sample was recruited via social media, which was likely the reason why our sample consisted of a higher proportion of females, higher level of education, and potentially higher levels of digital literacy than the general population [18]. A further limitation of this study was that we did not know the percentage of people who attended in-person consults during the study period, nor did we know whether those who did not access telehealth services required a health appointment or attended an in-person consult instead. In addition, our survey did not collect any information on the type of telehealth services that participants attended (eg, allied health and specialist health services). Future studies should investigate whether the factors associated with a poorer telehealth experience are similar across the models of health care delivery. Furthermore, future surveys should compare people’s experiences with telehealth to people’s experiences with traditional in-person visits based on the type of health service provided (eg, general practitioner, specialist, and allied health services) and determine the impact of different health service modalities and types on health outcomes.

Comparison With Prior Work

Our findings suggest that for some, telehealth was perceived as less effective for delivering mental health services. This is concerning, as mental health problems, such as depression and anxiety, were at least twice as prevalent during the first month of the COVID-19 pandemic in Australia compared to before the pandemic [19], and this problem is only expected to grow [20]. Our findings are also similar to previous research on telehealth and mental health [21], which is concerning, as negative experiences with telehealth may result in no mental health care for patients if face-to-face services are unavailable.

The most common theme for why respondents perceived telehealth to be worse than in-person medical care was less effective communication. This issue can be addressed by encouraging the use of established strategies for improving communication between health professionals and patients. For example, the teach-back method, which is also known as the show-me method or closing-the-loop method, has been shown to increase people’s understanding of health information by asking patients to repeat health information in their own words [22]. In addition, providing a written lay summary of the visit via a patient letter or patient portal may improve patients’ telehealth experiences. Patients have reported improved patient-provider communication as a result of using a patient portal [23]. Other approaches for addressing this issue may include online education or mobile apps, which have both been used to enhance patients’ understanding of content and improve health care services [24-27]. However, it should be noted that the Australian government is currently fast-tracking electronic prescribing, which may improve communication between general practitioners, patients, and pharmacists [28]. Electronic prescribing allows for the easy electronic sharing of prescriptions, thereby eliminating issues related to the challenge of obtaining prescriptions from telehealth appointments.

The following issues regarding access to telehealth services were also identified in our study: physical examination was not possible, people were less confident in their doctor or health care professional during telehealth visits, and additional burden was experienced for complex health conditions. These issues can be addressed by setting clear expectations for telehealth when scheduling appointments and explaining which types of appointments are suitable for telehealth. For example, when appointments are scheduled, patients should be notified that additional in-person consults may be required depending on the complexity of their medical appointment and that telehealth services may not be appropriate for mental health issues. In addition, videoconferencing could be offered, as it may allow for more reliable visual assessments and more accurate diagnoses [17]. Overall, in order to improve patients’ experiences with telehealth, strategies should be implemented to ensure that patients are aware of what to expect from telehealth appointments.

It is important to note that, in our study, 19.9% (272/1369) of respondents cancelled or postponed an in-person health appointment. The main reason for cancelling appointments was concern about catching COVID-19 at health or hospital clinics. Our study found a higher proportion of people who cancelled or postponed a health appointment than a survey conducted by the Australian Bureau of Statistics (11%) [5]. Similarly, a study of 151 women with breast cancer conducted in Israel found that 31% of people cancelled a health appointment, with the most common reason being the fear of contracting COVID-19 [29]. This is worrying, as continuing outbreaks may deter patients from accessing essential in-person medical care for some time. Therefore, our results suggest that telehealth services should continue to be offered while the community transmission of COVID-19 persists. Future studies should investigate whether patients who cancel or postpone health appointments are seeking telehealth services and monitor the long-term impact of the use of health services on health outcomes.

Future Directions and Considerations of Telehealth

It is worth noting that the temporary telehealth scheme is scheduled to end on March 31, 2021 [30]. It is unclear as to what degree and for whom telehealth services will be subsidized after this date. Telehealth has the potential to reduce inequality by making health care services more accessible. However, in order for telehealth to be an effective public health service, it should be widely available and affordable. Our findings showed participants' willingness to use telehealth services after the pandemic. However, we did not investigate whether people would engage with telehealth services if these services were no longer subsidized. It is reasonable to expect that people's willingness to use telehealth services may change, depending on funding reforms. Future research should continue to investigate patients' attitudes toward telehealth as policies change over time.

Our study provides insight into patients' perceptions of telehealth services. However, we did not investigate health professionals' perspectives and perceptions, including their experiences with telehealth, confidence to deliver telehealth, and willingness to continue to provide telehealth services beyond the pandemic. Further studies should investigate both patients' and health professionals' attitudes toward and experiences with telehealth, as they are both important voices in discussions about the future of telehealth in Australia.

Conclusions

Overall, we found that respondents' telehealth experiences were comparable to their experiences with traditional in-person visits. We identified the most common reasons for a poor experience with telehealth and provided strategies for improving the experience of telehealth users. In light of our results, telehealth may be worthwhile as a mode of health care delivery while the pandemic continues, and it may be worthwhile beyond the pandemic. However, studies on a broader, more representative sample of the Australian population are warranted.

Acknowledgments

We would like to acknowledge the members of the Australian public who participated in this survey.

Conflicts of Interest

None declared.

References

1. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
2. $1.1 billion to support more mental health, medicare and domestic violence services. Prime Minister of Australia. 2020 Mar 29. URL: https://www.pm.gov.au/media/11-billion-support-more-mental-health-medicare-and-domestic-violence-services-0 [accessed 2020-11-19]
3. Snoswell CL, Caffery LJ, Haydon HM, Thomas EE, Smith AC. Telehealth uptake in general practice as a result of the coronavirus (COVID-19) pandemic. Aust Health Rev 2020 Sep;44(5):737-740. [doi: 10.1071/AH20183] [Medline: 32853536]
4. Thomas EE, Haydon HM, Mehrotra A, Caffery LJ, Snoswell CL, Banbury A, et al. Building on the momentum: Sustaining telehealth beyond COVID-19. J Telemed Telecare 2020 Sep 26:1357633X20960638. [doi: 10.1177/1357633X20960638] [Medline: 32985380]
5. Household Impacts of COVID-19 Survey. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/people/people-and-communities/household-impacts-covid-19-survey/29-apr-4-may-2020 [accessed 2020-11-19]
6. Wilhelm JA, Helleringer S. Utilization of non-Ebola health care services during Ebola outbreaks: a systematic review and meta-analysis. J Glob Health 2019 Jun;9(1):010406 [FREE Full text] [doi: 10.7189/jogh.09.010406] [Medline: 30701070]
7. Wang SY, Chen LK, Hsu SH, Wang SC. Health care utilization and health outcomes: a population study of Taiwan. Health Policy Plan 2012 Oct;27(7):590-599. [doi: 10.1093/heapol/czr080] [Medline: 22258470]
8. Dow A. Warnings GP surgeries could close as pandemic hits doctor income. The Age. 2020 Sep 06. URL: https://www.theage.com.au/national/victoria/warnings-gp-surgeries-could-close-as-pandemic-hits-doctor-income-20200904-p55shq.html?utm_source=twitter&utm_medium=racgp&utm_campaign=47d567ce-66e6-4bf2-968c-47ac140e106e [accessed 2020-11-19]
9. McCaffery KJ, Dodd RH, Cvejic E, Ayre J, Batcup C, Isautier JMJ, et al. Health literacy and disparities in COVID-19–related knowledge, attitudes, beliefs and behaviours in Australia. Public Health Res Pract. Epub ahead of print 2020 Nov 05:e [FREE Full text] [doi: 10.17061/phrp30342012]

10. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
11. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
12. Wolf MS, Smith SG, Pandit AU, Condon DM, Curtis LM, Griffith J, et al. Development and Validation of the Consumer Health Activation Index. Med Decis Making 2018 Apr;38(3):334-343 [FREE Full text] [doi: 10.1177/0272989X17753392] [Medline: 29436308]
13. Socio-Economic Indexes for Areas (SEIFA) 2016. Australian Bureau of Statistics. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/2033.0.55.001Main+Features12016?OpenDocument [accessed 2020-11-19]
14. Weber R. Basic Content Analysis, Second Edition. Thousand Oaks, CA: SAGE Publications; 1990.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
16. Telehealth - Here to Stay? Global Centre for Modern Ageing. 2020 Jun. URL: https://static1.squarespace.com/static/5ae17ed32971146319f879ca/t/5f446c2012ef3956120b20df/1598319661535/Telehealth+Here+to+Stay+GCMA+research+report.pdf [accessed 2020-11-19]
17. Rush KL, Howlett L, Munro A, Burton L. Videoconference compared to telephone in healthcare delivery: A systematic review. Int J Med Inform 2018 Oct;118:44-53. [doi: 10.1016/j.ijmedinf.2018.07.007] [Medline: 30153920]
18. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interv 2016 May 27;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
19. Lam AY, Jorm AF, Wong DF. Mental health first aid training for the Chinese community in Melbourne, Australia: effects on knowledge about and attitudes toward people with mental illness. Int J Ment Health Syst 2010 Jun 24;4:18 [FREE Full text] [doi: 10.1186/1752-4458-4-18] [Medline: 20576137]
20. Parrish E. The next pandemic: COVID-19 mental health pandemic. Perspect Psychiatr Care 2020 Jul;56(3):485. [doi: 10.1111/ppc.12571] [Medline: 32602165]
21. Pellicano L, Brett S, de Houting J, Heyworth M, Magiati I, Steward R, et al. “I want to see my friends”, The everyday experiences of autistic people and their families during COVID-19. The University of Sydney. URL: https://www.sydney.edu.au/content/dam/corporate/documents/sydney-policy-lab/everyday-experiences-of-autistic-people-during-covid-19---report---july-2020.pdf [accessed 2020-11-19]
22. Barenfanger J, Sautter RL, Lang DL, Collins SM, Hacek DM, Peterson LR. Improving patient safety by repeating (read-back) telephone reports of critical information. Am J Clin Pathol 2004 Jun;121(6):801-803. [doi: 10.1309/9DYM-6R0T-M830-U95Q] [Medline: 15198350]
23. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
24. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
25. Wu X, Guo X, Zhang Z. The Efficacy of Mobile Phone Apps for Lifestyle Modification in Diabetes: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
26. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an Internet-Delivered Exercise and Pain-Coping Skills Training Intervention for Persons With Chronic Knee Pain: A Randomized Trial. Ann Intern Med 2017 Apr 04;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
27. De Oliveira Silva D, Pazzinatto MF, Crossley KM, Azevedo FM, Barton CJ. Novel Stepped Care Approach to Provide Education and Exercise Therapy for Patellofemoral Pain: Feasibility Study. J Med Internet Res 2020 Jul 22;22(7):e18584 [FREE Full text] [doi: 10.2196/18584] [Medline: 32706674]
28. Electronic Prescriptions - For Prescribers. Australian Government Australian Digital Health Agency. URL: https://www.digitalhealth.gov.au/get-started-with-digital-health/electronic-prescriptions/for-prescribers [accessed 2020-11-19]
29. Shinan-Altman S, Levkovich I, Tavori G. Healthcare utilization among breast cancer patients during the COVID-19 outbreak. Palliat Support Care 2020 Aug;18(4):385-391 [FREE Full text] [doi: 10.1017/S1478951520000516] [Medline: 32594966]
30. $2 billion to extend critical health services across Australia 2020. Prime Minister of Australia. 2020 Sep 18. URL: https://www.pm.gov.au/media/2-billion-extend-critical-health-services-across-australia [accessed 2020-11-19]

Edited by G Eysenbach; submitted 23.09.20; peer-reviewed by C Snoswell, A Jani, D De Oliveira Silva; comments to author 06.10.20; revised version received 25.10.20; accepted 31.10.20; published 10.12.20.

Please cite as:
Isautier JMJ, Copp T, Ayre J, Cvejic E, Meyerowitz-Katz G, Batcup C, Bonner C, Dodd R, Nickel B, Pickles K, Cornell S, Dakin T, McCaffery KJ
People's Experiences and Satisfaction With Telehealth During the COVID-19 Pandemic in Australia: Cross-Sectional Survey Study
J Med Internet Res 2020;22(12):e24531
URL: http://www.jmir.org/2020/12/e24531/
doi:10.2196/24531
PMID:33156806

Original Paper

Use of Telemedicine for Emergency Triage in an Independent Senior Living Community: Mixed Methods Study

Kelsi Carolan[1], MSW, PhD; David C Grabowski[2], PhD; Ateev Mehrotra[2], MD, MPH; Laura A Hatfield[2], PhD

[1]School of Social Work, University of Connecticut, Hartford, CT, United States
[2]Department of Health Care Policy, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Kelsi Carolan, MSW, PhD
School of Social Work
University of Connecticut
38 Prospect St
Hartford, CT, 06103
United States
Phone: 1 959 200 3680
Email: kelsi.carolan@uconn.edu

Abstract

Background: Older, chronically ill individuals in independent living communities are frequently transferred to the emergency department (ED) for acute issues that could be managed in lower-acuity settings. Triage via telemedicine could deter unnecessary ED transfers.

Objective: We examined the effectiveness of a telemedicine intervention for emergency triage in an independent living community.

Methods: In the intervention community, a 950-resident independent senior living community, when a resident called for help, emergency medical technician–trained staff could engage an emergency medicine physician via telemedicine to assist with management and triage. We compared trends in the proportion of calls resulting in transport to the ED (ie, primary outcome) in the intervention community to two control communities. Secondary outcomes were telemedicine use and posttransport disposition. Semistructured focus groups of residents and staff were conducted to examine attitudes toward the intervention. Qualitative data analysis used thematic analysis.

Results: Although the service was offered at no cost to residents, use was low and we found no evidence of fewer ED transfers. The key barrier to program use was resistance from frontline staff members, who did not view telemedicine triage as a valuable tool for emergency response, instead perceiving it as time-consuming and as undermining their independent judgment.

Conclusions: Engagement of, and acceptance by, frontline providers is a key consideration in using telemedicine triage to reduce unnecessary ED transfers.

(J Med Internet Res 2020;22(12):e23014) doi:10.2196/23014

KEYWORDS

telemedicine; telehealth; independent senior living communities; emergency care; first responders

Introduction

Nationwide, roughly one million people live in independent senior living communities [1], and residents often receive fragmented medical services. Senior living communities provide many amenities, but residents typically receive medical care from off-site physicians. When residents of senior living communities experience a new medical issue, the lack of an on-site physician may lead to unnecessary transfers to the emergency department (ED). Introducing on-site medical services within independent living communities may safely prevent unnecessary transfers [2,3] but may also be cost-prohibitive.

Telemedicine is a promising, less costly alternative for acute care in community settings [4]. In a randomized study in Massachusetts, United States, the introduction of off-hour telemedicine coverage in nursing homes decreased the hospital transfers by 11% [5]. However, the effectiveness of telemedicine for triage in independent senior living communities has never been evaluated. We studied the introduction of telemedicine for emergency triage at one such community. The goal of the program was to reduce the number of medical calls that resulted in transfers to the ED.

Methods

Overview

Telemedicine was introduced at one of three independent senior living communities in California, United States, managed by a single company. The intervention community was home to 950 residents. The remaining two communities, with similar staff models and residents, did not implement the program and served as comparators.

Clinical Services at Three Senior Living Communities

Safety staff in all three communities provided 24-hour response to resident calls for assistance, typically via a call pendant. Nearly all safety staff were certified as emergency medical technicians (EMTs). Residents of the communities were aware that the safety staff were trained for emergency medicine, as this was one of the selling points of living in such a community. As such, resident calls directly to the emergency telephone number 911 were rare. Upon arrival at a call, safety staff completed an initial assessment and made a triage decision between a life-threatening issue (ie, advanced life support [ALS] transport to the ED), an urgent issue (ie, basic life support [BLS] transport to the ED), less urgent care that still required same-day assessment (ie, transport to the ED via personal vehicle), or treat and release. Treat and release could involve follow-up in the following days with a personal physician. Obviously, residents indicated their wishes and in some cases decided against transfer to the ED against medical advice. For ED transfers, the safety staff member met the ambulance and facilitated the transfer. Following the call, staff documented the call in an electronic call log. Resident calls directly to 911 were not included in this study.

The three communities each had an on-site clinic staffed by nurses and nurses' aides during weekday hours. The on-site clinic focused primarily on medication delivery or minor issues such as a rash. These staff did not typically become involved in acute complaints.

Telemedicine Triage Intervention

After the introduction of telemedicine in May 2017, safety staff at the intervention community completed the same initial assessment; they then had the option of offering the resident a telemedicine consult. Safety staff were asked to utilize telemedicine for cases that were not clearly urgent and/or if there was uncertainty as to whether transport was warranted. If the safety staff felt the call was a life-threatening emergency, the telemedicine intervention was not utilized. If telemedicine was offered and the resident accepted, safety staff initialized a videoconference call on a tablet that they brought to all calls.

The telemedicine visits were provided by an emergency medicine physician who worked for a large national emergency medicine staffing company that had introduced the telemedicine service as a new care option. The visits were provided for free as part of this pilot project. There were a number of presentations about the service at regular facility meetings, including one where senior leadership from the telemedicine company attended a town hall meeting to promote awareness of the service and answer questions.

The responding emergency medicine physician worked remotely with the safety staff member to conduct an exam, using observations aided by stethoscope, blood pressure cuff, and pulse oximetry. The safety staff member, physician, and resident decided whether an ED transfer was warranted. The remaining response and follow-up procedures were the same.

Quantitative Analysis

Our quantitative analysis used deidentified medical call logs from the three communities. Variables included date; chief complaint; whether telemedicine was offered and accepted, if applicable; the outcome; and whether transported residents were admitted to the hospital. We excluded accidental activations and nonmedical calls. The format of the call logs differed among the three communities, requiring some reconciliation (see Multimedia Appendix 1). Our primary outcome was the proportion of calls resulting in residents being transported to the ED. Secondary outcomes were telemedicine use (ie, offers and acceptance) and disposition after transfer to the ED (ie, admitted to hospital or not).

Call log data were available from January 2017 to August 2018. In May 2017, telemedicine was introduced; in December 2017, a new policy required safety staff to offer telemedicine on all calls for which BLS transport would otherwise be called. Previously, this was at the discretion of the staff. Thus, we divided the data into three periods: the *Pre* period was from January to April 2017, the *Early* period was from May to November 2017, and the *Late* period was from December 2017 to August 2018. Quantitative analyses were performed in R v3.5.1 (The R Foundation).

Qualitative Analysis

Overview

Qualitative data can provide key insights into the decision-making processes of subjects [6], rendering it particularly useful for investigating facilitating factors and barriers in the adoption of novel medical interventions. Toward this aim, the authors conducted four semistructured focus group interviews in June 2018.

Participant Recruitment and Data Collection

Resident focus groups were advertised during the intervention facility's regularly held, facility-wide resident town halls and were further advertised through flyers posted throughout the facility. Two resident focus groups were conducted by the authors and included 19 residents with and without personal experience of telemedicine calls (see interview guides in Multimedia Appendix 2). Participating residents were asked to fill out anonymous demographic forms, but very few did so, rendering the demographic information collected insufficiently representative of the sample to report. Staff were recruited through the facility's resident director, who asked staff both on and off shift to consider participating and compensated staff for their time spent in a focus group. The two staff focus groups included 4 safety staff and 3 nurses' aides from the on-site clinic. Informed consent was obtained from all participants prior to commencing the focus group interviews and recording. The study was approved by the Institutional Review Board of the Longwood Medical Area in Boston, Massachusetts, United States.

Data Analysis

Interviews were recorded and professionally transcribed. The first author led qualitative data analysis, using a thematic analysis approach and NVivo 12 software (QSR International) to organize resulting themes. Employing a thematic analysis approach to analyzing the data allowed for the identification of themes, both within each unit of data analysis (ie, each transcribed focus group) and across the data set [7]. Thematic analysis is a six-step qualitative method of finding and interpreting patterns in the data: familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing a report [7]. The authors utilized a theory-driven approach to thematic analysis, which involves focusing on a previously identified research objective [7]; in the case of this study, the objective was to identify factors that facilitated or acted as barriers to resident participation in the intervention.

The first step of thematic analysis, familiarization with the data, began during data collection and preparation, with the authors discussing and taking notes between focus groups on initial ideas and potential patterns. This process continued through the first author's verification of the transcripts for accuracy, which involved multiple close readings of each transcript to confirm alignment with the audio recordings. The second step of thematic analysis, initial coding, involves beginning to identify and group data into relevant categories [7]. The first author looked for manifest and latent data, coding for content explicitly stated in the data, as well as attending to statements that highlighted a participant's attitude or assumptions [7].

After completing initial coding, the first author organized initial codes into related groups and began to search for preliminary themes within each of these lists of codes organized by topic [7]. At this juncture, the first author presented organized initial codes and resulting preliminary themes to the senior author for feedback, with the two authors discussing these preliminary results until a consensus on preliminary themes was reached. Next, qualitative transcripts were loaded into NVivo 12, and the first author utilized NVivo 12 to organize and review the initial themes identified through this consensus process. This review involved reconsidering both the relevance and importance of each initial theme as it related to the research objective and the congruence of identified themes across the data set as a whole, resulting in a further refined set of themes that were then named and defined [7]. The analysis process concluded with the first author presenting identified themes to all four authors, engaging in additional discussion and refinement of themes until consensus was reached. As all four authors participated in the data collection via co-leading focus groups, engaging in this final step supported the trustworthiness of findings by creating an opportunity to re-examine the extent to which the presented themes sufficiently represented the data and to build consensus on the salience of finalized themes.

Results

Quantitative Results

Average monthly call volumes rose over time in all three communities: from 67 (SD 6) to 80 (SD 11) in the intervention community, and from 42 (SD 11) to 46 (SD 13) and from 21 (SD 5) to 23 (SD 10) in the two comparison communities, respectively (see Figure 1). The average number of telemedicine calls per month was 4.2 (SD 1.9) in the *Early* period (5.5% of all calls, 25/456) and 5.3 (SD 3.5) in the *Late* period (6.6% of all calls, 53/805) (see Figure 2). The policy to mandate use of telemedicine for BLS calls increased the rate of refusals from 2.9% (13/456) to 16.0% (129/805) of calls.

At the intervention community, the fraction of calls resulting in ED transfer decreased from 52.4% (140/267) in the *Pre* period to 33.5% (270/805) in the *Late* period (ie, 18% decrease) (see Figure 3). At the comparison communities, calls decreased by 8% in one and increased by 8% in the other. To assess whether the change observed in the intervention community could be driven by growth of telemedicine, we compared the observed change in the ED transport rate at the intervention community to the theoretical change, in which every telemedicine call that did not result in ED transport had instead resulted in transport. This yields an upper bound on the possible impact of telemedicine. Of the 53 total telemedicine calls in the *Late* period, 37 (70%) were not transported. If, in the absence of telemedicine, all of these had been transported, the maximum possible effect of telemedicine would have been to decrease the transport rate by 5% versus by the observed rate of 18%.

If telemedicine prevents unnecessary ED transfers, the remaining transfers likely would be higher acuity and more likely to result in hospital admissions; however, we saw the opposite (see Figure 4). Over time, the percentage of transported calls that were admitted decreased from 42.1% (59/140) in the *Pre* period and 42.2% (86/204) in the *Early* period to 35.2% (95/270) in the *Late* period. Therefore, our evidence is insufficient to attribute the decrease in the transport rate at the intervention community to telemedicine.

Figure 1. Number of medical calls at the intervention and the two comparison communities. Dashed vertical lines indicate the demarcations between *Pre* and *Early* periods (ie, the start of telemedicine in May 2017) and between *Early* and *Late* periods (ie, the start of the new telemedicine use policy in December 2017).

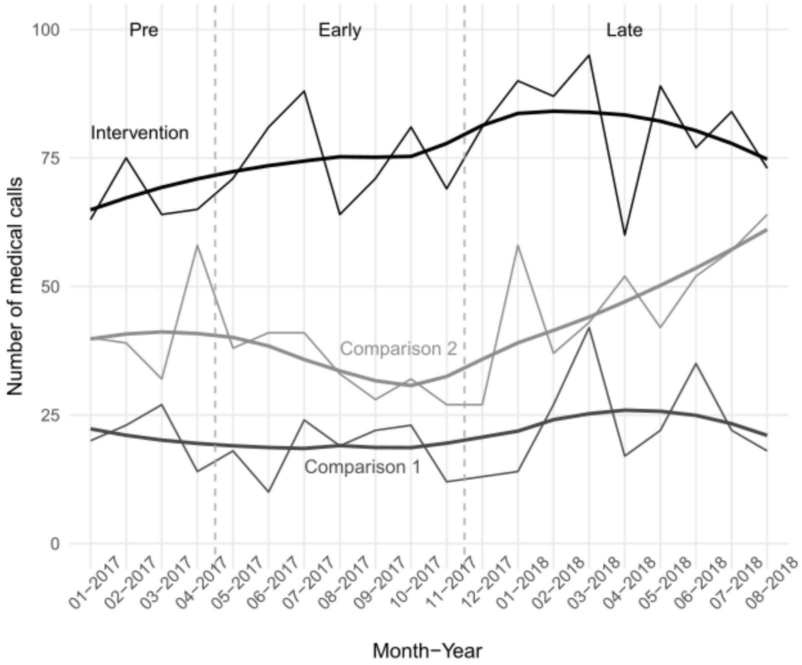


Figure 2. Number of telemedicine calls and refusals at the intervention community. Dashed vertical lines indicate the demarcations between *Pre* and *Early* (ie, the start of telemedicine in May 2017) and between *Early* and *Late* periods (ie, the start of the new telemedicine use policy in December 2017).

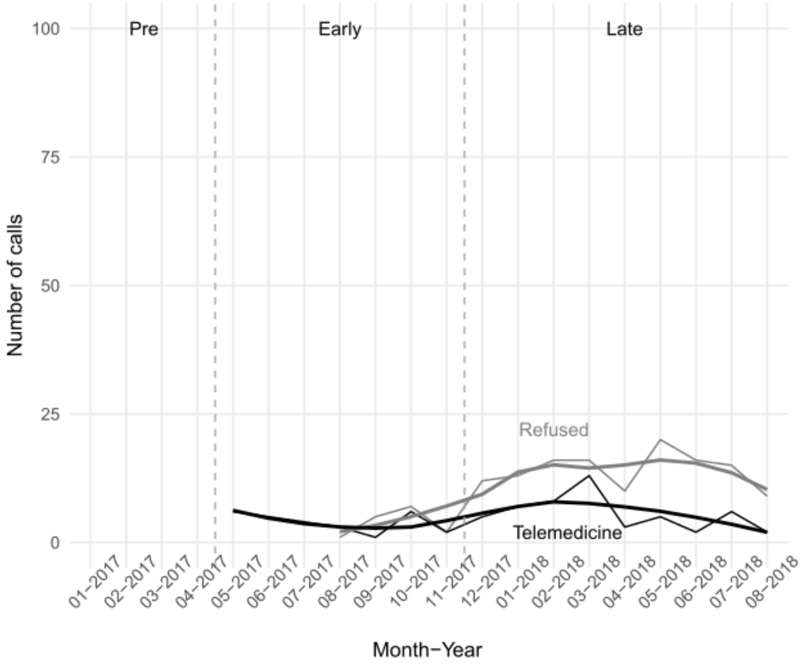


Figure 3. Percentage of calls resulting in transport to the emergency department by ambulance at the intervention and two comparison communities. Dashed vertical lines indicate the demarcations between *Pre* and *Early* (ie, the start of telemedicine in May 2017) and between *Early* and *Late* (ie, the start of the new telemedicine use policy in December 2017).

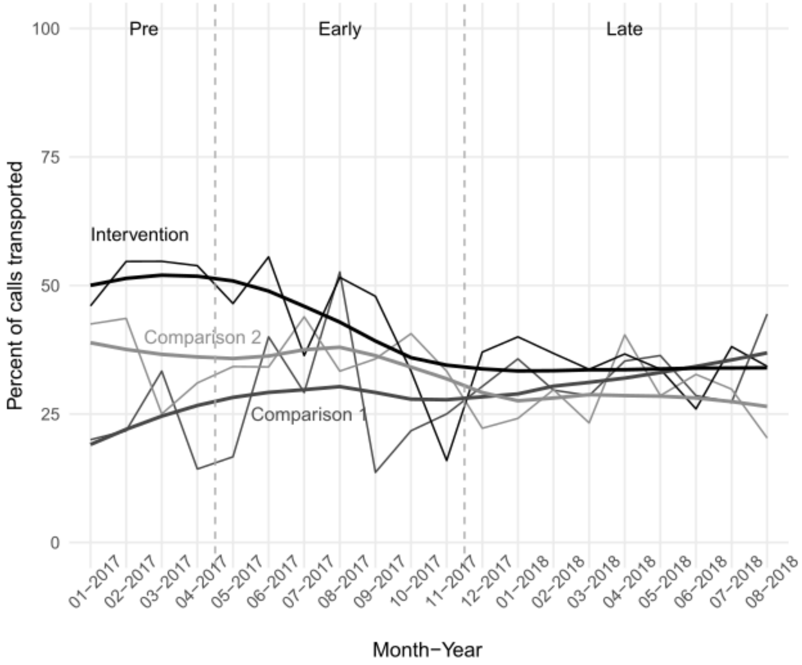
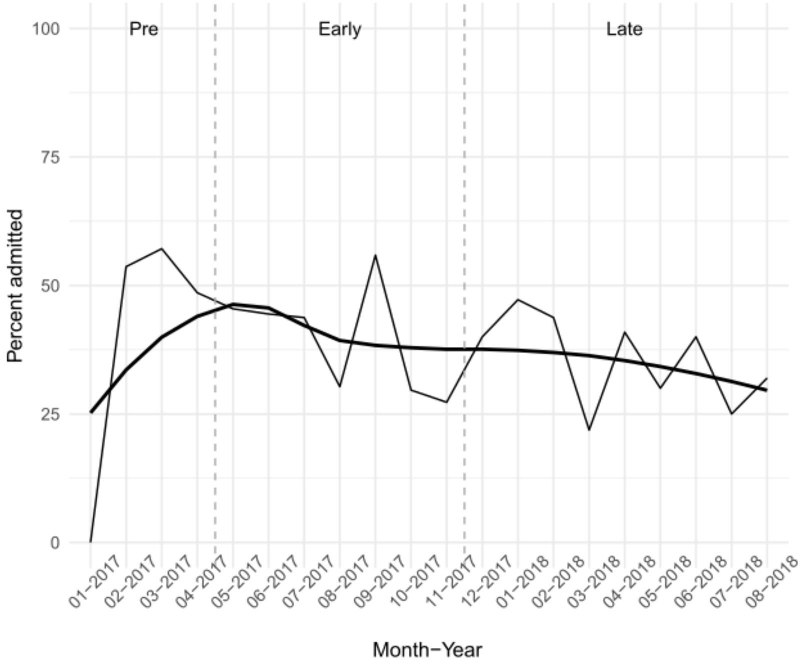


Figure 4. Percentage of transported calls from the intervention community that resulted in a hospital admission. Dashed vertical lines indicate the demarcations between *Pre* and *Early* (ie, the start of telemedicine in May 2017) and between *Early* and *Late* periods (ie, the start of the new telemedicine use policy in December 2017).



Qualitative Results: Staff Perspectives

Overview

EMT-trained safety staff did not perceive telemedicine as a valuable tool, viewing it as undermining their autonomy in decision making and increasing their workload. Reducing resident transports to the ED was not a goal embraced by safety staff, who described approaching all calls with the assumption that transport would likely be necessary.

Telemedicine: "Delaying the Inevitable"

Safety staff approached calls with the perspective that resident transport to the ED was likely: "We normally respond to all emergencies like probably they are going to consist of transports." Another safety staff participant reiterated, "...the majority of the calls that I get, they usually warrant transport anyway, so I just go, 'There's no point [in a telemedicine consult].' That's kind of where I'm at with it." If the medical need for transport appeared uncertain, staff preferred to call an ambulance: “If there is a borderline [case]...maybe you err on the side of just calling EMS [emergency medical services] out to handle that assessment.” Operating from the perspective that transport was likely necessary and/or best in the case of uncertainty, staff viewed telemedicine as an unnecessary obstacle, delaying the resident’s needed emergency transport: “...we have to wait for the telemedicine conference, it still ends up in transport anyway, so it’s just delaying the inevitable.” Thus, safety staff also perceived telemedicine as undermining their professional judgment:

> *That was my first [time using telemedicine] and she ended up going out anyway. The resident was...complaining of dizziness, couldn’t make it back to her unit from outside. The [telemedicine] doc was basically like, “Oh, just go rest. Check on her every 20 minutes.” Went and checked on her 20 minutes later. She had nausea and vomiting and was passed out on the floor, couldn’t keep herself up, so she still went out on ALS transport. That was one of them that...I was definitely frustrated, because I already thought she should go out ALS...That was, for me, I never really used it after that. I used it a couple of times, and each time it still ended up in the transport, too.*

This safety staff participant independently determined that transport was necessary, and regarded this decision as undermined by the telemedicine physician. When transport did end up being necessary in this case, the safety staff participant’s view of telemedicine consult as an obstacle rather than an aide was cemented.

Telemedicine Is Not a Good “Fit”

Staff expressed that telemedicine was of limited use for the community’s residents: “...the majority of stuff we deal with here, it doesn’t fit the bill. Old people don’t bounce. They fall and break.” Another participant described common conditions for which residents call safety staff: “...shortness of breath, and fevers...but most of the time, they’re 102, they’re septic, they’re coughing up phlegm, their lungs are nasty. You’re like, ‘They need to go.’” One participant indicated that falls with pain are frequent: “I always tell them, ‘I can’t see through you, so you’re going to have to go get an x-ray or an MRI [magnetic resonance imaging] or whatever it needs to be.’” Staff highlighted that they were accustomed to making independent triage decisions and that the telemedicine physician “asking the same questions” was not helpful.

Telemedicine for Nonemergency Use

Safety staff felt that the program’s goal to deter ED visits was the wrong goal and that telemedicine would better address minor medical concerns at the on-site clinic. Safety staff indicated that “the resistance you have from our department is we’re emergency response,” whose “primary goal is security, medical response.” The nursing staff at the on-site clinic “deal with the nonemergency, ‘Hey, I don’t feel well,’ and they walk into their office looking for the Tylenol.” Safety staff indicated more willingness to utilize the telemedicine consult if the safety staff called to the case had already determined for themselves that transport was not necessary. A participant indicated that telemedicine could be beneficial if and when “I don’t think it warrants going to the hospital. It may be an alternative to going to their own doctor or to an urgent care. If I don’t see emergency treatment as really necessary.” The safety staff’s self-description of their role as emergency responders and the emphasis on telemedicine as potentially more suitable for nonemergency cases highlights their perception of telemedicine intervention as inappropriate for emergency use.

Perceived Burden Without Perceived Benefit

Safety staff indicated that telemedicine increased their workload, describing the software as difficult to navigate: “Just the sheer amount of time that it takes...typically there’s three [safety staff on duty]. We have well over 900 residents we’re checking on. Sometimes the telemedicine can take a little longer than I think is warranted.” This was echoed by another staff member who described telemedicine as increasing the difficulty of juggling multiple demands: “It can be time-consuming. We have a lot of things going on, more than one medical going on...it can be very time-consuming.” Safety staff participants also described a number of technical issues that acted as a barrier to use, such as inconsistent wireless internet connectivity throughout the facility: “...that’s come up in certain parts of the campus...We just try and make do and move around if possible.” In combination with the safety staff’s perception of telemedicine as inappropriate for emergency situations, the staff’s experience of telemedicine as additional work led to the perception of telemedicine as adding burden without adding value.

Residents Are Reluctant

Safety staff also reported that residents were reluctant to use telemedicine. One safety staff member said, “Honestly, they don’t directly request it, and I think a lot of that has to do with they don’t want to change from the old-school days of actually seeing a doctor. A lot of them, they go, ‘No, they can’t do anything for me over the phone or over the tablet. I actually need to go in.’” Another staff member agreed, saying, “...when I do offer it to residents, they only want to see their doctor and then they want to be in person...to talk to someone who knows them personally.” One participant described this perceived reluctance as a lack of comfort with receiving medical care through an unfamiliar medium:

> *I think it’s a generational thing. I think once all of us are old, you hand me a tablet and go “Here’s the doctor,” I’m going to go, “Alright, let’s do it,” but...It was different back when they saw a doctor, they had that personal care, so a lot of them, I think that’s part of the problem. They’re like, “I just want to see my doctor.”*

Qualitative Results: Residents

Overview

In contrast to staff perceptions, resident focus group participants described multiple benefits of using telemedicine. Residents emphasized that avoiding an ambulance transport and having their concerns successfully resolved by the telemedicine physician were the most significant benefits of a telemedicine consult.

Avoiding the ED

Residents expressed a strong interest in avoiding trips to the ED whenever possible, mentioning the long wait times, financial costs, and potential health risks of ED visits. Residents interviewed identified avoiding an ED visit as a primary benefit of using the telemedicine intervention. A resident described satisfaction with the quality of the telemedicine consults she received, emphasizing that an unnecessary transfer to the ED was prevented in both of her experiences with the telemedicine intervention:

> *The doctor I spoke with was very understanding and he went through all the questions he should pertaining to my problem, and I was very reassured...he had my medical record right in front of him, and so he knew all of my problems and that also enlightened him as to how to treat me. The second time it was the same experience and the same thing. I had another doctor, and I was very pleased. In both instances it did save a trip going to the ER.*

Another resident emphasized the value of telemedicine in preventing emergency transport:

> *Well, I had this situation before, so I thought, "Okay, if this keeps me out of the hospital, good." And he did. He stayed on with me for a long time and after he left, everything was fine and I didn't need to go to the hospital...I think that's where telemedicine is a good thing, saving people going to the emergency room.*

Other residents who had not yet tried telemedicine indicated a willingness to do so if the opportunity arose. A resident expressed the following:

> *I think it's a fine service that would be very beneficial and, in many cases, prevent someone from going to an emergency room if they didn't need to. I'd rather have a medical opinion available to me on the spot to make that determination... I would not have the EMTs say, "Well, we're going to take you..." So I think it's a good deal.*

Reinforcing the Need for Emergency Transport

Residents also described an unanticipated benefit of telemedicine: encouraging hesitant residents to go to the emergency room when necessary. A resident described his experience as follows:

> *The skeptical patient is the one that this is directed to. I passed out after walking up a hill, and when I woke up I felt good...So I just thought, "Well, it's just passing out. I'm going to go ahead..." And then the telemedicine person said, "Well, assuming your recent history in hospitals, we would recommend that you go to the emergency room."*

The participant described how he agreed to the transport and was diagnosed with a "fatal heart condition" in the ED. He concluded, "Everybody I've talked to about telemedicine, I just tell them they saved my life...convincing me that it was necessary to go. I wouldn't have gone."

Resident Concerns: Delaying Treatment

Resident focus group data indicated that residents were satisfied with the results of telemedicine use and/or expressed a willingness to try it. A few residents did indicate concern that telemedicine would delay treatment:

> *My concern is it delays my journey to a physician. And sometimes you need a CT [computed tomography] scan, sometimes you need some really heavy meds, and if I have to spend 40 minutes at home going through telemed...that just slows the whole process down. So, to me, it's a negative...If you're in a lot of pain, you don't want to mess with telemedicine. You want to get somewhere where you can be treated right away.*

However, the majority of residents expressed confidence in their ability to determine when telemedicine might be useful.

Discussion

Principal Findings

Unnecessary ED visits are a major source of excess morbidity and spending [8] and are common in independent senior living communities. Telemedicine has the potential to help deter unnecessary ED visits [4]. In our evaluation of a telemedicine option in a senior living community, we did not observe a clear decrease in ED transfers due to the telemedicine intervention. The use of the tool was low, which appeared to be driven by staff opposition to the model. Residents viewed the service more favorably, with several asserting it prevented an unnecessary transfer.

Our findings echo other studies in this area where securing provider buy-in is an ongoing challenge for many organizations implementing telemedicine interventions [9]. In a study of a telemedicine intervention in several skilled nursing facilities, some nurses refused to use the service [5]. In evaluations of heart failure telemedicine interventions, medical personnel felt it reduced autonomy and increased workload [10]. The views of frontline providers are key to successful implementation of telemedicine interventions.

Minimal research has explored the feasibility and acceptability of telemedicine interventions involving first responders as frontline providers [11]. The intervention facility's senior management considered EMT-trained safety staff as an advantage, as staff would be able to effectively assist the remote physician. However, the safety staff viewed the intervention as undermining their autonomy. Traditionally, prehospital first responders are trained to provide minimal treatment to stabilize patients for emergency transport, with successful transfer to the ED as the goal [12]. EMTs embracing reduced ED transport as a goal would require a significant shift in professional mentality and culture.

Staff may have benefitted from further education on identifying potentially avoidable transfers, the harm to residents of unnecessary transfers, and the opportunity to work and learn in cooperation with remote physicians. Designating an individual frontline provider as a *champion* for the program may help to encourage telemedicine adoption by other frontline staff [9,13]. In this case, the champion was the director of the safety staff, who was not a frontline provider. Having a frontline provider as champion may have helped communicate staff concerns and helped educate the safety staff about the program.

Financing of telemedicine interventions in independent living settings presents an additional key challenge. This program was offered free of charge to residents living in the independent living community because the services were donated in kind by the telemedicine provider as part of the demonstration project. For telemedicine programs to be more widely adopted, they will need to be financed by residents or third-party payers. Payers like Medicare Advantage or accountable care organizations that are at risk for transfers to the ED appear willing to invest in these programs. There have already been clinical investments in nursing homes [14] and assisted living facilities [3,15] by Medicare Advantage plans. The Centers for Medicare and Medicaid Services recently released a new payment model reimbursing emergency ambulance service providers for telemedicine interventions provided to Medicare beneficiaries [16]. This may incentivize the increased use of telemedicine by ambulance service providers and increase interest in the adoption of telemedicine services in long-term care settings.

Limitations

Our study has several key limitations. It is based on the adoption of telemedicine at one senior living community, which may not generalize to other communities. The focus groups were convenience samples that may not be representative of the entire population of residents and staff. Residents of these communities, on average, were from high socioeconomic strata and are not representative of the general population of older adults in independent senior living communities. The two comparison communities used slightly different staffing models and call log data formats, limiting the comparability of the data. The number of telemedicine calls each month was very low, limiting statistical power. This study was conducted prior to the COVID-19 pandemic, during which the use of telemedicine has grown rapidly. A telemedicine triage intervention may now be more attractive to both frontline providers and residents.

Conclusions

We found that the introduction of telemedicine in an independent living community did not clearly reduce transfers to the ED. The intervention had low uptake due to limited buy-in by frontline staff. Future interventions involving telemedicine use in independent living communities should carefully consider how best to prepare and engage frontline providers.

Acknowledgments

This research was supported by funding from the Patrick and Catherine Weldon Donaghue Medical Research Foundation.

Conflicts of Interest

DG recently served on the Scientific Advisory Committee for NaviHealth. We have no further conflicts of interest to disclose.

Multimedia Appendix 1
Details of the call log variables.
[DOC File , 51 KB - jmir_v22i12e23014_app1.doc]

Multimedia Appendix 2
Semistructured interview guides.
[DOC File , 40 KB - jmir_v22i12e23014_app2.doc]

References

1. Freedman VA, Spillman BC. The residential continuum from home to nursing home: Size, characteristics and unmet needs of older adults. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69 Suppl 1:S42-S50 [FREE Full text] [doi: 10.1093/geronb/gbu120] [Medline: 25342822]
2. Bynum JP, Andrews A, Sharp S, McCollough D, Wennberg JE. Fewer hospitalizations result when primary care is highly integrated into a continuing care retirement community. Health Aff (Millwood) 2011 May;30(5):975-984 [FREE Full text] [doi: 10.1377/hlthaff.2010.1102] [Medline: 21555482]
3. Ptaszek AL, Kane RL. Evaluation of the Erickson Advantage Continuing Care Retirement Community Demonstration. Palo Alto, CA: Pacific Consulting Group; 2008 Nov. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.613.9297&rep=rep1&type=pdf [accessed 2020-12-04]
4. Hwang U, Shah MN, Han JH, Carpenter CR, Siu AL, Adams JG. Transforming emergency care for older adults. Health Aff (Millwood) 2013 Dec;32(12):2116-2121 [FREE Full text] [doi: 10.1377/hlthaff.2013.0670] [Medline: 24301394]
5. Grabowski DC, O'Malley AJ. Use of telemedicine can reduce hospitalizations of nursing home residents and generate savings for medicare. Health Aff (Millwood) 2014 Feb;33(2):244-250. [doi: 10.1377/hlthaff.2013.0922] [Medline: 24493767]
6. Dohan D, Garrett SB, Rendle KA, Halley M, Abramson C. The importance of integrating narrative into health care decision making. Health Aff (Millwood) 2016 Apr;35(4):720-725. [doi: 10.1377/hlthaff.2015.1373] [Medline: 27044974]

7. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
8. Institute of Medicine. Hospital-Based Emergency Care: At the Breaking Point. Washington, DC: The National Academies Press; 2007.
9. Ellimoottil C, An L, Moyer M, Sossong S, Hollander JE. Challenges and opportunities faced by large health systems implementing telehealth. Health Aff (Millwood) 2018 Dec;37(12):1955-1959. [doi: 10.1377/hlthaff.2018.05099] [Medline: 30633667]
10. Boyne JJJ, Vrijhoef HJM. Implementing telemonitoring in heart failure care: Barriers from the perspectives of patients, healthcare professionals and healthcare organizations. Curr Heart Fail Rep 2013 Sep;10(3):254-261. [doi: 10.1007/s11897-013-0140-1] [Medline: 23666901]
11. Rörtgen D, Bergrath S, Rossaint R, Beckers SK, Fischermann H, Na I, et al. Comparison of physician staffed emergency teams with paramedic teams assisted by telemedicine--A randomized, controlled simulation study. Resuscitation 2013 Jan;84(1):85-92. [doi: 10.1016/j.resuscitation.2012.06.012] [Medline: 22750663]
12. Nurok M, Henckes N. Between professional values and the social valuation of patients: The fluctuating economy of pre-hospital emergency work. Soc Sci Med 2009 Feb;68(3):504-510. [doi: 10.1016/j.socscimed.2008.11.001] [Medline: 19062149]
13. Ouslander J, Lamb G, Tappen R, Herndon L, Diaz S, Roos B, et al. Interventions to reduce hospitalizations from nursing homes: Evaluation of the INTERACT II collaborative quality improvement project. J Am Geriatr Soc 2011 Apr;59(4):745-753. [doi: 10.1111/j.1532-5415.2011.03333.x] [Medline: 21410447]
14. Kane RL, Flood S, Bershadsky B, Keckhafer G. Effect of an innovative medicare managed care program on the quality of care for nursing home residents. Gerontologist 2004 Feb;44(1):95-103. [doi: 10.1093/geront/44.1.95] [Medline: 14978325]
15. Grabowski DC, Caudry DJ, Dean KM, Stevenson DG. Integrated payment and delivery models offer opportunities and challenges for residential care facilities. Health Aff (Millwood) 2015 Oct;34(10):1650-1656 [FREE Full text] [doi: 10.1377/hlthaff.2015.0330] [Medline: 26438740]
16. Centers for Medicare and Medicaid Services. HHS launches innovative payment model with new treatment and transport options to more appropriately and effectively meet beneficiaries' emergency needs. US Department of Health & Human Services. 2019 Feb 14. URL: https://www.hhs.gov/about/news/2019/02/14/hhs-launches-innovative-payment-model-new-treatment-transport-options.html [accessed 2019-04-03]

Abbreviations

ALS: advanced life support
BLS: basic life support
CT: computed tomography
ED: emergency department
EMS: emergency medical services
EMT: emergency medical technician
MRI: magnetic resonance imaging

Edited by G Eysenbach; submitted 10.08.20; peer-reviewed by E Burner, S Sarbadhikari; comments to author 15.09.20; revised version received 10.11.20; accepted 14.11.20; published 17.12.20.

Please cite as:
Carolan K, Grabowski DC, Mehrotra A, Hatfield LA
Use of Telemedicine for Emergency Triage in an Independent Senior Living Community: Mixed Methods Study
J Med Internet Res 2020;22(12):e23014
URL: http://www.jmir.org/2020/12/e23014/
doi:10.2196/23014
PMID:33331827

Original Paper

Practice-Level Variation in Telemedicine Use in a Pediatric Primary Care Network During the COVID-19 Pandemic: Retrospective Analysis and Survey Study

Kelsey Schweiberger[1], MD; Alejandro Hoberman[1,2], MD; Jennifer Iagnemma[2], DNP, RN; Pamela Schoemer[2], MD; Joseph Squire[2], RN, BSN; Jill Taormina[2], RN, BSN; David Wolfson[2], MD; Kristin N Ray[1,2], MD, MS

[1]Department of Pediatrics, University of Pittsburgh School of Medicine, UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA, United States
[2]UPMC Children's Community Pediatrics, Pittsburgh, PA, United States

Corresponding Author:
Kelsey Schweiberger, MD
Department of Pediatrics
University of Pittsburgh School of Medicine
UPMC Children's Hospital of Pittsburgh
3414 Fifth Avenue
Pittsburgh, PA, 15213
United States
Phone: 1 412 692 6809
Email: schweibergerka@upmc.edu

Abstract

Background: Telehealth, the delivery of health care through telecommunication technology, has potential to address multiple health system concerns. Despite this potential, only 15% of pediatric primary care clinicians reported using telemedicine as of 2016, with the majority identifying inadequate payment for these services as the largest barrier to their adoption. The COVID-19 pandemic led to rapid changes in payment and regulations surrounding telehealth, enabling its integration into primary care pediatrics.

Objective: Due to limited use of telemedicine in primary care pediatrics prior to the COVID-19 pandemic, much is unknown about the role of telemedicine in pediatric primary care. To address this gap in knowledge, we examined the association between practice-level telemedicine use within a large pediatric primary care network and practice characteristics, telemedicine visit diagnoses, in-person visit volumes, child-level variations in telemedicine use, and clinician attitudes toward telemedicine.

Methods: We analyzed electronic health record data from 45 primary care practices and administered a clinician survey to practice clinicians. Practices were stratified into tertiles based on rates of telemedicine use (low, intermediate, high) per 1000 patients per week during a two-week period (April 19 to May 2, 2020). By practice tertile, we compared (1) practice characteristics, (2) telemedicine visit diagnoses, (3) rates of in-person visits to the office, urgent care, and the emergency department, (4) child-level variation in telemedicine use, and (5) clinician attitudes toward telemedicine across these practices.

Results: Across pediatric primary care practices, telemedicine visit rates ranged from 5 to 23 telemedicine visits per 1000 patients per week. Across all tertiles, the most frequent telemedicine visit diagnoses were mental health (28%-36% of visits) and dermatologic (15%-28%). Compared to low telemedicine use practices, high telemedicine use practices had fewer in-person office visits (10 vs 16 visits per 1000 patients per week, P=.005) but more total encounters overall (in-office and telemedicine: 28 vs 22 visits per 1000 patients per week, P=.006). Telemedicine use varied with child age, race and ethnicity, and recent preventive care; however, no significant interactions existed between these characteristics and practice-level telemedicine use. Finally, clinician attitudes regarding the usability and impact of telemedicine did not vary significantly across tertiles.

Conclusions: Across a network of pediatric practices, we identified significant practice-level variation in telemedicine use, with increased use associated with more varied telemedicine diagnoses, fewer in-person office visits, and increased overall primary care encounter volume. Thus, in the context of the pandemic, when underutilization of primary care was prevalent, higher practice-level telemedicine use supported pediatric primary care encounter volume closer to usual rates. Child-level telemedicine use differed by child age, race and ethnicity, and recent preventive care, building upon prior concerns about differences in access to telemedicine. However, increased practice-level use of telemedicine services was not associated with reduced or increased differences in use, suggesting that further work is needed to promote equitable access to primary care telemedicine.

(J Med Internet Res 2020;22(12):e24345) doi:10.2196/24345

KEYWORDS

telehealth; telemedicine; ambulatory pediatrics; health services research; ambulatory; pediatrics; health services; COVID-19

Introduction

Telehealth, the delivery of health care through telecommunication technology, has potential to address multiple health system concerns; it can alleviate physician workforce shortages, improve access to care, mitigate disparities in health care, control costs, and enhance communication between clinicians [1-3]. Despite this potential, the uptake of telehealth among pediatric clinicians has largely remained outside of primary care pediatrics, with growth instead observed in mental health, subspecialty care, and direct-to-consumer telemedicine provided by clinicians outside of the medical home [4,5]. The American Academy of Pediatrics (AAP) cautions against pediatric telemedicine provided outside of the primary care office due to concerns about fragmentation of care, suboptimal care quality, and lack of integrated follow-up; however, the AAP supports integration of telehealth into primary care pediatrics within the patient-centered medical home [1,6].

Despite the AAP endorsement of telehealth within primary care, only 15% of pediatric primary care clinicians reported using telemedicine as of 2016, with the majority identifying inadequate payment for these services as the largest barrier [7]. As of February 2020, all state Medicaid programs had payment provisions for live video telehealth services; however, only 19 states paid for telehealth services when the patient was located in their home, and only 5 states mandated payment parity with in-person visits [8]. This situation was reflected in similar stipulations by commercial payers. With limited payment options for telehealth services, especially for patients located at home, the adoption of telemedicine was not financially viable for most pediatric primary care offices outside of integrated care delivery systems before March 2020 [7,9].

In March 2020, the COVID-19 pandemic precipitated a rapid need for increased telehealth services to safely deliver care while limiting the risk of exposure to contagion that is inherent in an in-person setting [10]. The nationwide need for telehealth services led to rapid changes in payment and regulations surrounding telehealth delivery. Specifically, updated policies allowed patients to be located in their homes during a telemedicine visit and allowed use of widely available technology platforms to deliver telehealth by waiving penalties for Health Insurance Portability and Accountability Act (HIPAA) violations [11,12]. These policy changes, and the shifting perceptions of risk versus benefit of in-person and virtual care, enabled the sudden adoption of telemedicine within primary care practices across the country [13-16].

Thus, we are witnessing an acute surge in telemedicine use within pediatric primary care; however, much is unknown about the potential uses and impact of telehealth in pediatric primary care, given the prior rarity of this model of care. In this paper, we describe the experience of a large pediatric primary care network within the first two months of telemedicine use during the COVID-19 pandemic. Specifically, we aimed to identify high versus low telemedicine-using primary care practices and compare (1) practice characteristics, (2) telemedicine visit diagnoses, (3) rates of in-person visits to the office, urgent care, and the emergency department (ED), (4) child-level variation in telemedicine use, and (5) clinician attitudes toward telemedicine across these practices.

Methods

Context and Study Population

We performed a retrospective analysis of electronic health record (EHR) data from 45 practices within a large pediatric primary care network. These practices are certified as patient-centered medical homes by The Joint Commission; together, they provide care for approximately 315,000 children throughout Western Pennsylvania across 13 counties. All the practices shared one EHR, which offered embedded video visits through a patient portal. Some of these practices had briefly trialed a model of acute care telemedicine in 2015; however, none of the practices were offering telemedicine services at the start of the pandemic.

Telemedicine Implementation

Local payers began to offer payment for telemedicine when the patient is at home on March 17, 2020 [17]. On March 23, 2020, the first county-specific stay-at-home order was issued in Pennsylvania [17,18]. Several practice leads trialed multiple telemedicine platforms and workflows from March 18-20, 2020, with implementation strategies shared with all practices via videoconference on March 23, 2020. The network quality and safety leaders led collaborative learning videoconferences two to three times per week for the next two months with all physicians, advanced practice providers (APPs), and practice managers, sharing telemedicine best practices and discussing other COVID-19–related topics. Initial relaxation of the stay-at-home order for the largest metropolitan county occurred on May 15, 2020, with transition of the state to its “yellow phase.”

Data Source

We obtained encounter data for all telemedicine and in-person visits between March 18 and May 2, 2020, from the EHR, and we identified patient panels for each practice, defined as all patients with one or more encounters at the practice in the prior two years. For each practice’s patient panel, patient demographics, date of last preventive visit, and counts of telemedicine, office, ED, and urgent care visits were obtained from the EHR. To complement this EHR-based evaluation, we also surveyed the primary care clinicians across the practice network regarding the usability, usefulness, and perception of patient and clinician experience of telemedicine.

EHR Data and Variables

For each practice in the network, the number of practice clinicians (doctors and APPs) and practice site locations were obtained from network records. Practice site counties were classified as rural or urban using the 2013 rural-urban continuum codes [19].

Telemedicine visits were identified using EHR encounter type codes. All completed telemedicine visits from March 18 to May 2, 2020, were included. For each telemedicine visit, we extracted the practice site, age of the child on the date of visit, and primary visit diagnosis.

Telemedicine visit primary diagnoses were categorized based on International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) diagnosis codes into 22 broad ICD-10-CM diagnostic categories corresponding to organ systems [20]. For nonspecific categories (eg, “symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified” (R00-R99)), we reviewed the subcategories and recategorized them into the relevant organ system. For example, the subcategory of “symptoms and signs involving the skin and subcutaneous tissue” (R20-R23) was grouped with “diseases of the skin and subcutaneous tissues” (L00-L99) into “skin and subcutaneous tissue diagnoses.”

To compare the volumes of telemedicine visits with other modes of care delivery, we extracted each practice’s volumes from the EHR, including in-person office, urgent care, and ED visits, as well as telephone encounters (excluding those with a telemedicine or in-person visit on the same day) during a 2-week window during the pandemic (April 19 to May 2, 2020).

Finally, for each child identified as part of a practice’s panel, we extracted age, race and ethnicity, health insurance (Medicaid vs commercial), and whether the child had a preventive visit within the prior 12 months. The child’s race and ethnicity were originally recorded in the EHR based on parent response during the child’s first visit with the practice. Across all practices, 82% of patients were identified as White non-Hispanic; therefore, analyses by patient race and ethnicity were limited to comparing children identified as White non-Hispanic to children identified as any other racial or ethnic identity, of which the majority identified as “other” (9%), non-Hispanic Black (8%), or Hispanic (1%).

Clinician Survey

To complement this primarily EHR-based analysis, clinicians in the primary care network were invited to participate in a web-based survey. The survey items examined the usability and usefulness of telemedicine through items modified from the Technology Usability Questionnaire, which is a validated survey incorporating questions from the Technology Acceptance Model, Telemedicine Satisfaction Questionnaire, and Post-Study System Usability Questionnaire and encompasses five subscales: usefulness, ease of use, effectiveness, reliability, and satisfaction, with all items using Likert scales [21]. We added questions pertaining to the physician experience of telemedicine, including perceived impact on quality of care (informed by the Institute of Medicine’s six domains of quality), impact on job satisfaction (informed by self-determination theory), and perceived usefulness of telemedicine for different visit reasons [22,23]. The survey included 37 questions, and it is available in Multimedia Appendix 1. Participants had the option to identify their practice or leave this item blank.

Clinicians were invited to participate in the anonymous web-based survey from April 28 to May 14, 2020, through a series of 4 emails. The timeframe was chosen to capture summative attitudes and experiences of clinicians coinciding with the end of the EHR-based analysis.

Identification of Low, Intermediate, and High Telemedicine Use Practices

For each practice, to account for variation in practice size, we determined the total number of telemedicine visits completed per week and divided the number of visits by the number of active patients in that practice to provide a standardized rate of telemedicine visits per 1000 patients. To categorize high versus low telemedicine use practices, the rate of telemedicine visits per 1000 patients per week was averaged for the 2-week period (April 19 to May 2, 2020) occurring after the first month of telemedicine implementation. This time frame was chosen to categorize practices at a time where telemedicine visit volume had stabilized so that the analysis could focus on practices with high versus low use (as opposed to early vs late adopters). The 45 practices were categorized into tertiles based on their telemedicine visit rates, which we labeled as low, intermediate, or high telemedicine use practices.

Statistical Analysis

The analyses compared practice-level characteristics, telemedicine visit diagnoses, in-person visit volumes, variation in volume by patient characteristics, and clinician attitudes across telemedicine use tertiles. Across practice-level telemedicine use tertiles, we compared practice-level characteristics and in-person visits per 1000 patients per week using Kruskal-Wallis tests. We compared the percentage of telemedicine visits within each individual diagnosis category across tertiles using logistic regression.

For patients in each practice’s panel, we determined the percentage of children who had one or more telemedicine visits during the 2-week period of the focused analysis (April 19 to May 2, 2020) by specific child characteristics (age, race and ethnicity, insurance type, and receipt of a preventive visit in the prior year). First, to assess whether there was a significant difference in telemedicine use by child characteristics across all practices (regardless of practice-level telemedicine use), we used a child-level logistic regression model clustered by practice but with no tertile variable. To determine whether increased practice-level use of telemedicine altered the differences in telemedicine use by child characteristics, we assessed the significance of an interaction term between each child characteristic category and practice tertile within a series of child-level logistic regression models clustered by practice across all tertiles. We tested the significance of the interaction terms using Wald tests, and we present the results as adjusted percentages determined through these models.

Clinician survey responses were compared across tertiles using linear regression clustered by practice, excluding respondents who did not identify their practice. We also compared responses for respondents who did not identify their practice versus those who did, again using linear regression, and found no significant differences.

All analyses were conducted in Stata version 16.1 (StataCorp) with significance assessed using an alpha level of .05.

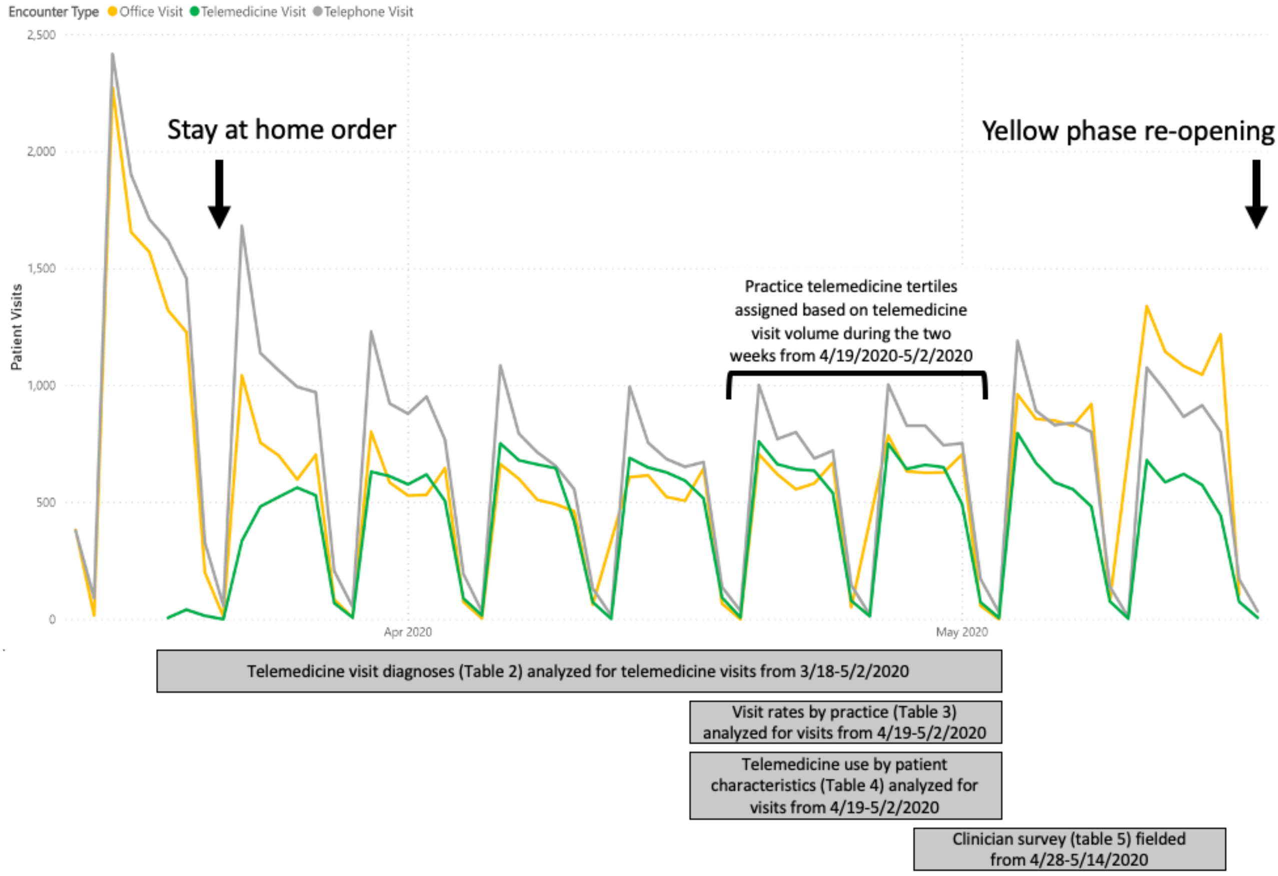
Approval and Ethical Considerations

This analysis was part of a quality improvement project aimed at improving pediatric primary care telemedicine delivery and was approved by the University of Pittsburgh Medical Center's Quality Review Committee. Projects approved by this committee do not meet the definition of human subjects research and therefore do not require formal approval by an institutional review board.

Results

Starting on March 23, 2020, the network underwent a rapid reduction of in-person office visit volumes and a simultaneous increase in telemedicine visit volumes (Figure 1). Telemedicine visit volume reached a steady state three to four weeks after implementation.

Figure 1. Encounters by telemedicine, in-person office, and care via telephone across 45 practices within a pediatric primary care network, March through May 2020. Telemedicine use tertiles were defined based on visit volume from April 19 to May 2, 2020 (indicated by brackets), as this period represented steady-state telemedicine use. Yellow phase indicates the first relaxation of the stay-at-home order, indicating the first phase of reopening in Pennsylvania.



Practice-Level Characteristics

High telemedicine use practices had more physicians in the practice (median 4) than low telemedicine use practices (median 3, *P*=.04); however, all other practice-level characteristics were similar (Table 1).

Table 1. Practice characteristics by low, intermediate, or high telemedicine use (N=45).

Variable	Telemedicine use			*P* value
	Low	Intermediate	High	
Practices, n (%)	15 (33)	15 (33)	15 (33)	N/A[a]
Telemedicine visits per 1000 patients per week, range	5-9.6	9.7-14	15-23	N/A
Practice characteristics				
Active patients in practice, range	1200-8300	2600-7800	2000-10400	.08
Practice clinicians				
Physicians, median (IQR)	2 (1-6)	5 (2-10)	6 (4-10)	.04
APPs[b], median (IQR)	2 (1-3)	2 (2-6)	2 (1-5)	.70
Percent of patient population insured by Medicaid, median (IQR)	61 (50-77)	62 (53-73)	79 (62-82)	.07
Practice site (rural/urban)[c]				
Rural county, n (%)	4 (27)	4 (27)	3 (20)	.90

[a] N/A: not applicable.

[b] APPs: advanced practice providers.

[c] Based on US Department of Agriculture rural-urban continuum codes.

Telemedicine Visit Diagnoses

Across all tertiles, telemedicine visits were most common for mental health and skin/soft tissue–related diagnoses (Table 2). However, the percentage of visits in each of these diagnostic categories varied with practice-level telemedicine use. Visits with skin-related diagnoses, for example, comprised 17.5% of telemedicine visits (467/2661) at low telemedicine use practices, and 15.0% of visits (1435/9587) at high telemedicine use practices (P=.006). Although visits for skin-related diagnoses represented a smaller percentage of telemedicine visits at high telemedicine use practices, the number of skin-related visits per 1000 patients per week was higher at high telemedicine use practices compared to low telemedicine use practices (median 14 vs 7 telemedicine skin related visits per 1000 patients per week at high and low telemedicine use practices, respectively; P<.001). High telemedicine use practices had larger percentages of telemedicine visits devoted to respiratory (P<.001), ear/mastoid (P<.001), and genitourinary (P=.02) diagnoses than low telemedicine use practices (Table 2).

Table 2. Telemedicine visit diagnoses in practices (N=45) with low, intermediate, and high telemedicine use from March 18 to May 2, 2020.

Variable	Telemedicine use			*P* value[a]
	Low	Intermediate	High	
Telemedicine visits per 1000 patients per week, range	5-9.6	9.7-14	15-23	N/A[b]
Number of practices, n (%)	15 (33.3)	15 (33.3)	15 (33.3)	N/A
Visit diagnosis category[c] (total visits), n (%)				
Mental, behavioral, and neurodevelopmental diagnoses	969 (36.4)	1704 (31.9)	2724 (28.4)	<.001
Skin and subcutaneous tissue diagnoses	467 (17.5)	801 (15.0)	1435 (15.0)	.006
Respiratory system diagnoses	297 (11.2)	684 (12.8)	1395 (14.6)	<.001
Digestive system diagnoses	182 (6.8)	396 (7.4)	697 (7.3)	.60
Infectious and parasitic diagnoses	176 (6.6)	390 (7.3)	746 (7.8)	.04
Injury/poisoning	92 (4)	184 (3.4)	437 (4.6)	.001
General symptoms and signs	77 (3.5)	241 (4.5)	416 (4.3)	.01
Eye diagnoses	77 (3.5)	147 (2.8)	255 (2.7)	.50
Ear and mastoid process diagnoses	45 (1.7)	230 (4.3)	395 (4.1)	<.001
Genitourinary system diagnoses	45 (1.7)	123 (2.3)	240 (2.5)	.02
Other[d]	234 (8.8)	435 (8.2)	847 (8.8)	.60

[a] *P* values indicate significance of proportion of telemedicine visits within each individual diagnosis category compared across practice-level telemedicine use tertiles using logistic regression.

[b] N/A: not applicable.

[c] Categories based on codes in the International Classification of Diseases, Tenth Revision, Clinical Modification.

[d] Diagnostic categories with less than 2% of visits are represented in this "other" category; these included neoplasms and hematologic diagnoses, endocrine, nutritional and metabolic diseases, nervous system diagnoses, circulatory system diagnoses, musculoskeletal diagnoses , peripartum and perinatal diagnoses, congenital anomalies, and symptoms and signs not otherwise classified, as well as codes for special purposes, injuries, and contact with health services.

In-Person Visit Volume

Compared to low telemedicine use practices, practices with high telemedicine use had fewer in-person office visits (median 10 vs 16 in-person office visits per 1000 patients per week at high vs low telemedicine use practices, respectively; *P*=.005; Table 3). Practices with high telemedicine use also had slightly more ED visits (median 2 vs 1 ED visits per 1000 patients per week at high and low telemedicine use practices, respectively; *P*=.02) but similar urgent care visits. When accounting for both in-person and telemedicine office visits, high telemedicine use practices had more total primary care encounters per 1000 patients per week (median 28 vs 22 encounters, *P*=.006). Of note, among high telemedicine use practices, total primary care encounters (in-person and telemedicine) represented a median of 53% of the weekly volume in 2019. In contrast, among low telemedicine use practices, total primary care encounters represented a median of 46% of the 2019 weekly volume. Telephone encounters occurring separate from a visit were similar across tertiles.

Table 3. In-person visits by patients in practices with low, intermediate, or high telemedicine use (April 19 to May 2, 2020).

	Telemedicine use			*P* value[a]
	Low	Intermediate	High	
Practices, n (%)	15 (33)	15 (33)	15 (33)	N/A[b]
Primary care patient visits				
Telemedicine visits per 1000 patients per week, range	5-9.6	9.7-14	15-23	N/A
In-person office visits per 1000 patients per week, median (IQR)	16 (12-18)	11 (7-14)	10 (8-12)	.005
All primary care encounters per 1000 patients per week, median (IQR)	22 (19-26)	23 (19-26)	28 (25-30)	.006
In-person patient visits outside of primary care, median (IQR)				
Urgent care visits per 1000 patients per week	0.4 (0-1)	1 (0.5-2)	1 (0.8-2)	.10
Emergency department visits per 1000 patients per week	1 (0.8-1)	2 (1-2)	2 (1-2)	.02
Total encounters per 1000 patients per week outside of primary care	2 (1-3)	3 (2-4)	3 (2-4)	.008
Total encounters per 1000 patients per week	25 (20-28)	25 (24-29)	30 (28-33)	.003
Patient telephone calls to practice not associated with a visit, median (IQR)				
Telephone management without a visit per 1000 patients per week	17 (8-22)	14 (10-16)	15 (7-20)	.80

[a] *P* values indicate practice-level in-person visits per 1000 patients per week compared across practice-level telemedicine use tertiles using Kruskal-Wallis tests.

[b] N/A: not applicable.

Patient Characteristics Associated With Telemedicine Use

In logistic regression models without practice-level tertile designation, use of telemedicine varied significantly by child race and ethnicity (*P*<.001), child age (*P*<.001), and receipt of preventive care in the prior 12 months (*P*<.001) but did not vary by child insurance category (*P*=.40). However, the interaction terms between the practice tertile and each of these characteristics were not significant (Table 4). This indicates that although differences exist in the full sample for telemedicine use by child race and ethnicity, child age, and recency of a preventive visit, increasing practice-level use of telemedicine did not change the relationship between these patient characteristics and the likelihood of a telemedicine visit.

Table 4. Variation in telemedicine visits by child characteristic across practices with low, intermediate, and high telemedicine use. Adjusted percentages of children in patient panels who had one or more telemedicine visits during the study period using logistic regression clustered by practice. Italicized variables are significant in the entire population. *P* values indicate the significance of the interaction term between the specified characteristic and practice tertile.

Model with no interaction term		Model with tertile and telemedicine use interaction term			
	All practices	Low telemedicine use practices	Intermediate telemedicine use practices	High telemedicine use practices	*P* for interaction term
Practices, n (%)	45 (100)	15 (33.3)	15 (33.3)	15 (33.3)	N/A[a]
Telemedicine visits per 1000 patients per week, range	5-23	5-9.6	9.7-14	15-23	N/A
Children, n (%)	244,473 (100)	66,295 (27.1)	84,093 (34.4)	94,085 (38.5)	N/A
Adjusted percentage of children with one or more telemedicine visit, % (95% CI)					
By child age (years)[b]					.25
0-2	2.3 (2-2.7)	1.6 (0.8-2.3)	2.2 (1.9-2.6)	3.1 (2.6-3.5)	
3-5	1.3 (1.1-1.5)	1 (0.7-1.4)	1 (0.9-1.2)	1.7 (1.4-2)	
6-11	1.6 (1.5-1.8)	1.2 (0.9-1.4)	1.6 (1.5-1.8)	2 (1.8-2.2)	
12-17	1.6 (1.4-1.7)	1.1 (0.9-1.3)	1.6 (1.3-1.9)	1.9 (1.7-2)	
By race/ethnicity					.10
Non-Hispanic Black, Hispanic, and other race and ethnicity categories	1.4 (1.2-1.6)	0.7 (0.3-1.1)	1.3 (1.2-1.5)	1.8 (1.6-2.1)	
Non-Hispanic White	1.7 (1.6-1.9)	1.2(1-1.5)	1.7 (1.5-1.8)	2.1 (2-2.3)	
By receipt of preventive care within the last year					.26
Received preventive care in the last year	1.8 (1.7-2)	1.3 (1-1.6)	1.7 (1.6-1.9)	2.3 (2.1-2.5)	
Did not receive preventive care in the last year	1.1 (0.9-1.2)	0.7 (0.6-0.8)	1.1 (1-1.2)	1.3 (1.1-1.5)	
By insurance type[c]					.2
Medicaid-insured children	1.7 (1.5-1.9)	1.2 (0.9-1.6)	1.6 (1.4-1.8)	2.3 (2-2.6)	
Commercially insured children	1.6 (1.5-1.8)	1.1 (0.8-1.4)	1.6 (1.5-1.7)	2 (1.8-2.2)	

[a] N/A: not applicable.

[b] Patients over 18 years of age were excluded (n=20,424).

[c] Children whose insurance type was identified as self-pay were excluded (n=9258).

Attitudes of Clinicians Toward Telemedicine

The survey was completed by 121 clinicians, including 88 who identified their practice and 33 who did not (34% response rate). Clinician attitudes regarding the usability, usefulness for child health, usefulness for clinician experience, or impact of telemedicine did not vary significantly across tertiles (Table 5).

Table 5. Clinician perceptions of usability and usefulness of telemedicine.

Variable	Clinicians at low telemedicine use practices	Clinicians at intermediate telemedicine use practices	Clinicians at high telemedicine use practices	*P* value[a]
Telemedicine visits per 1000 patients per week, range	5-9.6	9.7-14	15-23	N/A[b]
Clinicians, n (%)	21 (24)	35 (40)	32 (36)	N/A
Practices represented	12	14	13	N/A
Usability, mean (SD)[c]				
Ease of use and learnability	5.8 (0.8)	5.4 (1.2)	5.9 (0.9)	.50
Effectiveness	4.9 (0.8)	4.4 (1.4)	4.5 (1.3)	.30
Reliability	2.3 (1.4)	2.5 (1.6)	2.2 (1.3)	.80
Satisfaction and future use	5 (1.1)	4.8 (1.7)	4.8 (1.3)	.70
Usefulness—child and population health, mean (SD)[d]				
Timeliness of care	3.8 (0.7)	3.7 (1)	3.8 (0.9)	.98
Equity in access to care	4 (0.8)	3.9 (1)	3.8 (1.1)	.50
Family-centeredness of care	3.4 (0.7)	3 (1)	3.1 (1.1)	.30
Health of my patients	2.8 (0.7)	3 (0.9)	2.9 (0.9)	.90
Continuity of care	3 (0.9)	3.1 (1.2)	3.1 (1)	.80
Safety of my patients	2.6 (0.7)	3.1 (1.1)	3 (1)	.20
Usefulness—clinician experience, mean (SD)[d]				
Financial health of my practice	2.8 (0.9)	2.8 (1)	2.6 (1.1)	.30
Sense of accomplishment from my work	2.7 (0.8)	2.6 (1)	2.2 (1)	.07
Satisfaction with how I spend my time	2.6 (1.1)	2.5 (1.1)	2.2 (1)	.20
Sense of connectedness with patients	2.3 (0.9)	2.6 (1.2)	2.3 (1.2)	.80
Suitability of telemedicine for specific reasons, mean (SD)[e]				
Acute care	2.2 (0.5)	2.3 (0.5)	2.3 (0.5)	.70
Chronic disease management	2.6 (0.7)	2.6 (0.7)	2.5 (0.7)	.70
Preventive care	1.9 (0.9)	1.8 (0.8)	1.8 (0.8)	.80
Follow-up care	2.4 (0.5)	2.3 (0.7)	2.3 (0.6)	.80
Care coordination	3.1 (0.8)	3 (0.7)	2.8 (0.7)	.30
Mental health	3.2 (0.7)	3 (0.5)	3 (0.7)	.30

[a] *P* values reported from clinician survey responses compared across tertiles using linear regression clustered by practice.

[b] N/A: not applicable.

[c] Survey questions answered using a 7-point Likert scale, where 1 indicates “strongly disagree” and 7 indicates “strongly agree.”

[d] Survey questions answered using a 5-point Likert scale, where 1 indicates “much worse,” 3 indicates “about the same,” and 5 indicates “much better.”

[e] Survey questions answered using a 4-point Likert scale, where 1 indicates “never” and 4 indicates “always.”

Discussion

Principal Findings

Among a network of pediatric practices, we examined practice-level variation in use of telemedicine during the COVID-19 pandemic, allowing us to explore ongoing questions about the relationship between telemedicine use, receipt of care, and equity in receipt of care. We found that increased practice-level telemedicine use was associated with more physicians in the practice, more varied telemedicine encounter diagnoses, and fewer in-person office visits per 1000 patients per week.

While concerns exist that telemedicine may promote overutilization, we found that in the context of the pandemic, greater pediatric primary care practice-level telemedicine use did not result in overutilization but rather supported primary care encounter volume slightly closer to usual rates during a time when underutilization of primary care and other health care settings was prevalent [24-28]. We also found that high telemedicine use practices had slightly more ED visits than low telemedicine use practices during the study period. Again, this finding must be interpreted in the setting of overall decreased health care encounters during this time, such that this slight increase could represent either a greater ability of telemedicine to triage children to emergent care during the COVID-19 pandemic or, alternatively, a slightly greater need for emergent care for patients in these practices. Another possibility is that high telemedicine use practices may have adhered more strictly to guidelines to reduce in-person office visits; that goal, rather than telemedicine itself, may have led to the slight increase in ED visits.

In addition to concerns about overutilization, concerns have been raised about whether increased telemedicine use will improve equity of access to care or exacerbate disparities in access due to the digital divide [2,29,30]. In our sample, children without a preventive visit in the past year and children who identified as races or ethnicities other than White non-Hispanic were less likely to have had a telemedicine visit, while telemedicine use did not vary by child insurance type. These differences in the overall population use of telemedicine build upon prior concerns about differences in access to broadband internet, patient portals, and telemedicine [29,31-34]. However, the nonsignificant interaction terms indicate that increased practice-level use of telemedicine services during this specific period neither reduced nor increased these differences. In cases where telemedicine is intended to be a tool to improve equity in access to care, these results indicate that simply increasing the use of telemedicine may not be sufficient to ensure more equitable access. Indeed, telemedicine has the potential to exchange one set of barriers to care (transportation issues, time constraints, hidden costs of missing work) for another (need for internet access, device capability, computer literacy). For telemedicine to more effectively reduce disparities in access, it will be necessary to implement telemedicine in ways that more intentionally overcome barriers.

While increased use of telemedicine by practices did not translate into greater equity in telemedicine use for patients, it did translate into more varied telemedicine use based on visit diagnoses. Additionally, we observed an increase in the proportion of visits related to respiratory and ear, nose, and throat (ENT) symptoms. Concerns have been raised about assessments of ears in the absence of tele-otoscopy, which was not available for the studied visits [35]. Given that clinicians in low versus high telemedicine use tertiles reported similar views about the suitability of telemedicine for acute care, increased respiratory and ENT diagnoses at high telemedicine use practices may reflect a stronger practice-wide commitment to reducing in-person visits due to the COVID-19 pandemic as opposed to increased comfort with caring for these diagnoses via telemedicine. For all practices, regardless of their use tertile, the most common telemedicine visit diagnosis categories were mental health and skin-related diagnoses. This differs from the most common diagnoses for visits by children to commercial direct-to-consumer telemedicine, where the most common visit diagnosis category was nose/sinus infections [4]. In contrast, the most common diagnosis group among these primary care telemedicine visits was mental health, which accounted for approximately 30% of telemedicine visits across the full set of practices.

Limitations

One key limitation of this EHR analysis is we cannot account for patient or parent preferences for or satisfaction with telemedicine use. Additionally, the racial and ethnic diversity within our sample was minimal, limiting analyses by race and ethnicity to a comparison of White non-Hispanic children and children identified as any other race or ethnicity. We also sought to compare visit volume by parental language; however, limited numbers of children with parental preference for non-English languages were identified in the relevant EHR field (0.5%). Although systems are in place to integrate ED and urgent care visit information into the EHR, ED and urgent care visits outside of our health system may still have been missed. However, we do not anticipate that this would result in any systematic bias across the practice tertiles given the large number of integrated ED and urgent care centers across the region. We also note that we focused this analysis on telemedicine visit diagnoses and volumes and did not examine quality of care or clinical outcomes. Finally, this analysis focused on telemedicine use within a specific context; therefore, the generalizability of the findings will need to be examined through other sources.

Conclusion

This study demonstrates that a large pediatric primary care network rapidly integrated use of telemedicine when given a favorable payment environment and public health mandates. The integration of telemedicine allowed continued contact with patients during the pandemic, largely for mental health care, with high practice-level telemedicine use allowing for more encounters with patients per week during a time where underutilization of primary care was common. Further work is needed to understand the sustainability of the pandemic-related surge in primary care telemedicine use and to identify ways to enhance the ability of telemedicine to improve access for those with access barriers.

Acknowledgments

KS is a postdoctoral scholar supported in part by a Health Resources and Services Administration NRSA for Primary Care Research Award (T32 HP22240). The funder had no role in the study design, analysis of data, manuscript preparation, or decision to submit for publication.

Authors' Contributions

KS and KNR conceptualized and designed the study, completed all statistical analyses, drafted the initial manuscript, and reviewed and revised the manuscript. AH, DW, PS, and JI conceptualized the study, coordinated the data collection, and reviewed and revised the manuscript. JT and JS collected data and reviewed and revised the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clinician survey.
[DOCX File , 19 KB - jmir_v22i12e24345_app1.docx]

References

1. Committee on Pediatric Workforce, Marcin JP, Rimsza ME, Moskowitz WB. The Use of Telemedicine to Address Access and Physician Workforce Shortages. Pediatrics 2015 Jul 29;136(1):202-209. [doi: 10.1542/peds.2015-1253] [Medline: 26122802]
2. Marcin JP, Shaikh U, Steinhorn RH. Addressing health disparities in rural communities using telehealth. Pediatr Res 2016 Jan 14;79(1-2):169-176. [doi: 10.1038/pr.2015.192] [Medline: 26466080]
3. NEJM Catalyst. What is telehealth? NEJM Catalyst. 2018 Feb 01. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0268 [accessed 2020-12-10]
4. Ray KN, Shi Z, Poon SJ, Uscher-Pines L, Mehrotra A. Use of Commercial Direct-to-Consumer Telemedicine by Children. Acad Pediatr 2019 Aug;19(6):665-669 [FREE Full text] [doi: 10.1016/j.acap.2018.11.016] [Medline: 30639759]
5. Barnett ML, Ray KN, Souza J, Mehrotra A. Trends in Telemedicine Use in a Large Commercially Insured Population, 2005-2017. JAMA 2018 Nov 27;320(20):2147-2149 [FREE Full text] [doi: 10.1001/jama.2018.12354] [Medline: 30480716]
6. Conners G, Kressly SJ, Perrin JM, Richerson JE, Sankrithi UM, Committee on Practice and Ambulatory Medicine, Committee on Pediatric Emergency Medicine, Section on Telehealth Care, Section on Emergency Medicine, Subcommittee on Urgent Care, Task Force on Pediatric Practice Change. Nonemergency Acute Care: When It's Not the Medical Home. Pediatrics 2017 May;139(5):e20170629 [FREE Full text] [doi: 10.1542/peds.2017-0629] [Medline: 28557775]
7. Sisk B, Alexander J, Bodnar C, Curfman A, Garber K, McSwain SD, et al. Pediatrician Attitudes Toward and Experiences With Telehealth Use: Results From a National Survey. Acad Pediatr 2020 Jul;20(5):628-635 [FREE Full text] [doi: 10.1016/j.acap.2020.05.004] [Medline: 32437881]
8. State Telehealth Laws and Reimbursement Policies: A Comprehensive Scan of the 50 States and District of Columbia. Center for Connected Health Policy. URL: https://www.cchpca.org/sites/default/files/2020-10/CCHP%2050%20STATE%20REPORT%20FALL%202020%20FINAL.pdf [accessed 2020-12-10]
9. Reed ME, Parikh R, Huang J, Ballard DW, Barr I, Wargon C. Real-Time Patient–Provider Video Telemedicine Integrated with Clinical Care. N Engl J Med 2018 Oct 11;379(15):1478-1479. [doi: 10.1056/nejmc1805746]
10. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care’s Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/nejmp2005835]
11. Notification of Enforcement Discretion for Telehealth Remote Communications During the COVID-19 Nationwide Public Health Emergency. US Department of Health and Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/notification-enforcement-discretion-telehealth/index.html [accessed 2020-12-10]
12. Medicare Telemedicine Health Care Provider Fact Sheet. Centers for Medicare & Medicaid Services. 2020 Mar 17. URL: https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet [accessed 2020-06-22]
13. Mehrotra A, Ray L, Brockmeyer DM, Barnett ML, Bender JA. Rapidly Converting to “Virtual Practices”: Outpatient Care in the Era of Covid-19. NEJM Catalyst. 2020 Apr 01. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0091 [accessed 2020-12-10]
14. Mueller B. Telemedicine Arrives in the U.K.: ‘10 Years of Change in One Week’. The New York Times. 2020 Apr 04. URL: https://www.nytimes.com/2020/04/04/world/europe/telemedicine-uk-coronavirus.html [accessed 2020-12-10]
15. Barney A, Buckelew S, Mesheriakova V, Raymond-Flesch M. The COVID-19 Pandemic and Rapid Implementation of Adolescent and Young Adult Telemedicine: Challenges and Opportunities for Innovation. J Adolesc Health 2020 Aug;67(2):164-171 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.006] [Medline: 32410810]
16. Wood SM, White K, Peebles R, Pickel J, Alausa M, Mehringer J, et al. Outcomes of a Rapid Adolescent Telehealth Scale-Up During the COVID-19 Pandemic. J Adolesc Health 2020 Aug;67(2):172-178 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.025] [Medline: 32611509]
17. Telemedicine Guidelines Related to COVID-19. Pennsylvania Department of Human Services. 2020 Mar. URL: https://www.dhs.pa.gov/providers/Quick-Tips/Documents/PROMISeQuickTip229.pdf [accessed 2020-07-21]
18. Order of the Governor of the Commonwealth of Pennsylvania for individuals to stay at home. Commonwealth of Pennsylvania Office of the Governor. 2020 Mar 23. URL: https://www.governor.pa.gov/wp-content/uploads/2020/03/03.23.20-TWW-COVID-19-Stay-at-Home-Order.pdf [accessed 2020-12-10]

19. Rural-Urban Continuum Codes. United States Department of Agriculture. 2013. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx [accessed 2020-12-10]
20. ICD-10-CM Official Guidelines for Coding and Reporting FY2020. Centers for Medicare & Medicaid Services. URL: https://www.cdc.gov/nchs/data/icd/10cmguidelines-FY2020_final.pdf [accessed 2020-12-10]
21. Parmanto B, Lewis AL, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
22. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; Aug 30, 2001:702-703.
23. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]
24. Santoli J, Lindley MC, DeSilva MB, Kharbanda EO, Daley MF, Galloway L, et al. Effects of the COVID-19 Pandemic on Routine Pediatric Vaccine Ordering and Administration - United States, 2020. MMWR Morb Mortal Wkly Rep 2020 May 15;69(19):591-593 [FREE Full text] [doi: 10.15585/mmwr.mm6919e2] [Medline: 32407298]
25. Hartnett KP, Kite-Powell A, DeVies J, Coletta MA, Boehmer TK, Adjemian J, National Syndromic Surveillance Program Community of Practice. Impact of the COVID-19 Pandemic on Emergency Department Visits - United States, January 1, 2019-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 12;69(23):699-704 [FREE Full text] [doi: 10.15585/mmwr.mm6923e1] [Medline: 32525856]
26. Mehrotra A, Chernew M, Linetsky D, Hatch H, Cutler D. The Impact of the COVID-19 Pandemic on Outpatient Visits: A Rebound Emerges. The Commonwealth Fund. 2020 May 19. URL: https://www.commonwealthfund.org/publications/2020/apr/impact-covid-19-outpatient-visits [accessed 2020-12-10]
27. Garcia S, Albaghdadi MS, Meraj PM, Schmidt C, Garberich R, Jaffer FA, et al. Reduction in ST-Segment Elevation Cardiac Catheterization Laboratory Activations in the United States During COVID-19 Pandemic. J Am Coll Cardiol 2020 Jun 09;75(22):2871-2872. [doi: 10.1016/j.jacc.2020.04.011] [Medline: 32283124]
28. Uchino K, Kolikonda MK, Brown D, Kovi S, Collins D, Khawaja Z, et al. Decline in Stroke Presentations During COVID-19 Surge. Stroke 2020 Aug;51(8):2544-2547. [doi: 10.1161/strokeaha.120.030331]
29. Velasquez D, Mehrotra A. Ensuring The Growth Of Telehealth During COVID-19 Does Not Exacerbate Disparities In Care. Health Affairs. URL: https://www.healthaffairs.org/do/10.1377/hblog20200505.591306/full/ [accessed 2020-12-10]
30. López L, Green AR, Tan-McGrory A, King RS, Betancourt JR. Bridging the Digital Divide in Health Care: The Role of Health Information Technology in Addressing Racial and Ethnic Disparities. The Joint Commission Journal on Quality and Patient Safety 2011 Oct;37(10):437-445. [doi: 10.1016/s1553-7250(11)37055-9]
31. Wilcock AD, Rose S, Busch AB, Huskamp HA, Uscher-Pines L, Landon B, et al. Association Between Broadband Internet Availability and Telemedicine Use. JAMA Intern Med 2019 Jul 29;179(11):1580 [FREE Full text] [doi: 10.1001/jamainternmed.2019.2234] [Medline: 31355849]
32. Walker DM, Hefner JL, Fareed N, Huerta TR, McAlearney AS. Exploring the Digital Divide: Age and Race Disparities in Use of an Inpatient Portal. Telemed J E Health 2020 May;26(5):603-613. [doi: 10.1089/tmj.2019.0065] [Medline: 31313977]
33. Drake C, Zhang Y, Chaiyachati KH, Polsky D. The Limitations of Poor Broadband Internet Access for Telemedicine Use in Rural America: An Observational Study. Ann Intern Med 2019 May 12;171(5):382. [doi: 10.7326/m19-0283]
34. Yoon H, Jang Y, Vaughan PW, Garcia M. Older Adults' Internet Use for Health Information: Digital Divide by Race/Ethnicity and Socioeconomic Status. J Appl Gerontol 2020 Jan 16;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]
35. Gerber JS. Need an Antibiotic? There's an App for That. Pediatrics 2019 May 08;143(5):e20190631 [FREE Full text] [doi: 10.1542/peds.2019-0631] [Medline: 30962254]

Abbreviations

AAP: American Academy of Pediatrics
APP: advanced practice provider
ED: emergency department
EHR: electronic health record
ENT: ear, nose, and throat
HIPAA: Health Insurance Portability and Accountability Act
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification

Edited by G Eysenbach; submitted 19.09.20; peer-reviewed by C Jacob, E van der Velde; comments to author 27.11.20; revised version received 02.12.20; accepted 03.12.20; published 18.12.20.

Please cite as:
Schweiberger K, Hoberman A, Iagnemma J, Schoemer P, Squire J, Taormina J, Wolfson D, Ray KN
Practice-Level Variation in Telemedicine Use in a Pediatric Primary Care Network During the COVID-19 Pandemic: Retrospective Analysis and Survey Study
J Med Internet Res 2020;22(12):e24345
URL: http://www.jmir.org/2020/12/e24345/
doi:10.2196/24345
PMID:33290244

Original Paper

Patient Interaction Phenotypes With an Automated Remote Hypertension Monitoring Program and Their Association With Blood Pressure Control: Observational Study

Anahita Davoudi[1*], MS, PhD; Natalie S Lee[2,3,4*], MPH, MSHP, MD; Corey Chivers[5], PhD; Timothy Delaney[6], BA; Elizabeth L Asch[3], BS; Catherine Reitz[6,7], MPH; Shivan J Mehta[6,7], MBA, MD; Krisda H Chaiyachati[3,6,7*], MPH, MSHP, MD; Danielle L Mowery[1,8*], MS, PhD

[1]Department of Biostatistics, Epidemiology & Informatics, University of Pennsylvania, Philadelphia, PA, United States
[2]National Clinician Scholars Program, University of Pennsylvania, Philadelphia, PA, United States
[3]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States
[4]Corporal Michael J Crescenz Veterans Affairs Medical Center, Philadelphia, PA, United States
[5]Penn Medicine Predictive Healthcare, University of Pennsylvania Health System, Philadelphia, PA, United States
[6]Center for Healthcare Innovation, University of Pennsylvania, Philadelphia, PA, United States
[7]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[8]Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, United States
[*]these authors contributed equally

Corresponding Author:
Natalie S Lee, MPH, MSHP, MD
National Clinician Scholars Program
University of Pennsylvania
423 Guardian Drive
13th Floor, Blockley Hall
Philadelphia, PA
United States
Phone: 1 215 898 6237
Email: natalie.lee@pennmedicine.upenn.edu

Abstract

Background: Automated texting platforms have emerged as a tool to facilitate communication between patients and health care providers with variable effects on achieving target blood pressure (BP). Understanding differences in the way patients interact with these communication platforms can inform their use and design for hypertension management.

Objective: Our primary aim was to explore the unique phenotypes of patient interactions with an automated text messaging platform for BP monitoring. Our secondary aim was to estimate associations between interaction phenotypes and BP control.

Methods: This study was a secondary analysis of data from a randomized controlled trial for adults with poorly controlled hypertension. A total of 201 patients with established primary care were assigned to the automated texting platform; messages exchanged throughout the 4-month program were analyzed. We used the *k*-means clustering algorithm to characterize two different interaction phenotypes: program conformity and engagement style. First, we identified unique clusters signifying differences in program conformity based on the frequency over time of error alerts, which were generated to patients when they deviated from the requested text message format (eg, ###/## for BP). Second, we explored overall engagement styles, defined by error alerts and responsiveness to text prompts, unprompted messages, and word count averages. Finally, we applied the chi-square test to identify associations between each interaction phenotype and achieving the target BP.

Results: We observed 3 categories of program conformity based on their frequency of error alerts: those who immediately and consistently submitted texts without system errors (perfect users, 51/201), those who did so after an initial learning period (adaptive users, 66/201), and those who consistently submitted messages generating errors to the platform (nonadaptive users, 38/201). Next, we observed 3 categories of engagement style: the enthusiast, who tended to submit unprompted messages with high word counts (17/155); the student, who inconsistently engaged (35/155); and the minimalist, who engaged only when prompted

(103/155). Of all 6 phenotypes, we observed a statistically significant association between patients demonstrating the minimalist communication style (high adherence, few unprompted messages, limited information sharing) and achieving target BP (P<.001).

Conclusions: We identified unique interaction phenotypes among patients engaging with an automated text message platform for remote BP monitoring. Only the minimalist communication style was associated with achieving target BP. Identifying and understanding interaction phenotypes may be useful for tailoring future automated texting interactions and designing future interventions to achieve better BP control.

(J Med Internet Res 2020;22(12):e22493) doi:10.2196/22493

KEYWORDS

text messaging; hypertension; telemedicine; cluster analysis

Introduction

Hypertension is a salient risk factor for heart disease and stroke [1]. Home blood pressure (BP) monitoring has long been accepted as a valid strategy for effective hypertension management [2]. Patients adhere to telemonitoring programs and use this technology for chronic conditions such as hypertension regardless of their nationality, socioeconomic status, or age [3]. In light of recent Centers for Medicare & Medicaid Services codes that reimburse remote monitoring of physiologic parameters such as BP and a shift toward greater remote monitoring because of the COVID-19 pandemic, remote BP monitoring is poised to become an increasingly common strategy for hypertension management.

An estimated 96% of Americans own mobile phones with text messaging capabilities [4], and mobile phone-based interventions are increasingly popular for remote disease management. Texting might be the best phone-based modality for disease management, as general patient populations are less likely to use smartphones, tablets, and health-related apps relative to texting [5]. Text messaging is an appealing platform for remote management given its accessibility and low costs, which may help reduce disparities in health care [6].

Automated texting in particular has been leveraged nationally and globally for remote hypertension management, with high patient engagement and satisfaction among low-income and underserved populations [7,8]. Automated texting interventions can double the odds of medication adherence for chronic conditions, regardless of whether texts are unidirectional or interactive [9]. Automated interactive texting may be an especially effective method of engaging patients in BP self-monitoring [10,11]. However, the impact of automated texting on BP targets is less clear, with mixed evidence about its effectiveness. Randomized trials suggest no effect of automated texting on achieving target BP [12,13], while other evidence suggests varying degrees of benefit [14,15].

These heterogeneous findings may be explained by different ways patients engage with texting platforms (behavioral phenotypes) [16]. Outside of changing the frequency of reminders or word choices, automated interventions using texting platforms have generally taken a one-size-fits-all approach to how patients should engage with the platform. However, patients likely differ in their interactions with and responses to automated texts. These variations are reflective of how individuals converse via text messaging (eg, quantity and quality of texting) and may correlate with how they engage in treatment, ultimately affecting clinical outcomes from automated texting interventions. We hypothesized that discrete behavioral phenotypes existed among patients engaging with clinical automated texting programs and that some phenotypes were likely to achieve targeted clinical outcomes (eg, controlled BP). Identifying and understanding the various ways patients engage with automated texting (phenotypes) would provide greater insights for targeting specific behaviors and tailoring interventions to improve hypertension control. For example, phenotypes associated with poor BP control may require more intensive texting approaches or more in-person care.

The purpose of this study was to identify and describe unique phenotypes of patient interactions with automated texting for remote BP management and estimate associations between interaction phenotypes and achieving a target BP.

Methods

Original Randomized Trial Description

In this University of Pennsylvania institute review board–approved study (828417, 834667), we retrospectively studied text messages and clinical data from adult patients in a randomized controlled trial (ClinicalTrials.gov NCT03416283) who were receiving automated text-based reminders for hypertension management. This trial's primary aim was to leverage automated text messaging for remote BP monitoring with or without social support to improve hypertension control over a 4-month program. Our study is a posttrial analysis that was independent of the study aims. Eligible participants were aged 18 to 75 years and had had at least 2 office visits at the Penn Family Care practice in Philadelphia, PA, with at least 2 office visit BP readings above goal (140/90 mm Hg) within 24 months prior to enrolling in the trial.

Study Context: Remote Monitoring With and Without Social Support Groups Trial

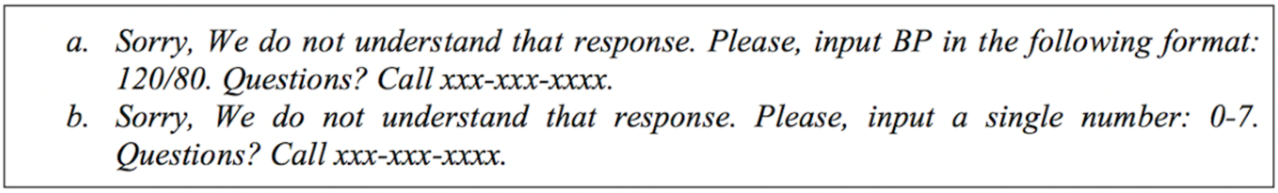
In the original study, scheduled automated text messages were used to (1) monitor each patient's BP measurements over time, (2) provide intermittent encouragement for engaging behavior, and (3) monitor BP medication adherence. Texts were sent through Way to Health, a Health Insurance Portability and Accountability Act–compliant, bidirectional, automated text communication platform used to engage with patients for research studies and clinical care delivery [17,18]. Patients in the original study received a variety of outbound texts including information about the study and upcoming in-person study visits. The three most frequent outbound messages include:

- Blood pressure prompts: "What is your blood pressure today (Ex. 120/80)?"
- Feedback: "You submitted X out of 3 BP measurements this week, great job!"
- Medication adherence: "How many days did you take your blood pressure medication(s) in the past 7 days? (Please input a single number: 0-7)"

The software was programmed to receive responses in a prespecified, structured format. For example, the reminder "What is your blood pressure today?" accepted a text response formatted as ###/## with few allowable variations, and the medication adherence prompt required a single number between 0 and 7. When patients submitted a correctly formatted text in response to these prompts, they received an automated confirmation text. However, as this was a text-based program, patients were not prevented from submitting text messages of any length in any form, at any time. Those message that did not conform to the requested text format triggered automated error messages. For example, when the text was not submitted in the expected format, patients received an automated error message with instructions to resubmit a response in the correct format. Not all formatting errors triggered an alert. For example, text messages sent immediately following an error message or messages that were not in direct response to an automated prompt (ie, were unprompted) did not generate these alerts. Examples of the reminder-specific error messages are shown in Figure 1.

There were two intervention arms. The first arm received automated text messages as described above. The second received the same automated text messages and identified a person in their social support network (eg, a family member or friend) who also received text-based reports about the study participant's performance in the program. The control arm received no text messages. Randomization was in a 2:2:1 ratio for the two intervention arms and a control group, with 201 patients randomized to the two text intervention arms. In the original trial, 101 patients were enrolled in the remote monitoring only group (RM) and 100 in the remote monitoring + social support group (RM+SS). BP readings that were consistently out of range were escalated to the clinical care team via the electronic health record.

Figure 1. Error messages: (a) blood pressure and (b) medication adherence.



a. Sorry, We do not understand that response. Please, input BP in the following format: 120/80. Questions? Call xxx-xxx-xxxx.

b. Sorry, We do not understand that response. Please, input a single number: 0-7. Questions? Call xxx-xxx-xxxx.

Characterizing Interaction Phenotypes

For our posttrial analysis, we focused on the 201 patients in the intervention arms to identify phenotypes of patient interactions with the automated text messaging system. We prespecified two categories of interaction phenotypes, program conformity and engagement style.

Program Conformity

Program conformity refers to the ways that users complied with the platform's requirements for text message communication. We used *k*-means clustering, an unsupervised learning method and data-driven approach to classify subgroups of observations within a dataset into *k* clusters based on each observation's proximity to a cluster mean or centroid [19]. Clusters were formed based on variations in error messages over time signifying program conformity. To capture the temporal variation of the error messages, we tallied the number of error messages at 20-day intervals and divided each 20-day total by the total number of messages by the patient in that time frame. Each patient must have submitted at least 1 message to be counted within a time frame and consistently engaged by submitting messages throughout all time frames to be included in the analysis. These features were fed into the *k*-means clustering algorithm to draw 3 clusters from the data signifying different program conformity user categories.

Engagement Style

Engagement style is characterized by observable patterns of engagement with the overall BP monitoring program. Engagement style patterns were based on a broader family of variables that were prespecified, including the following:

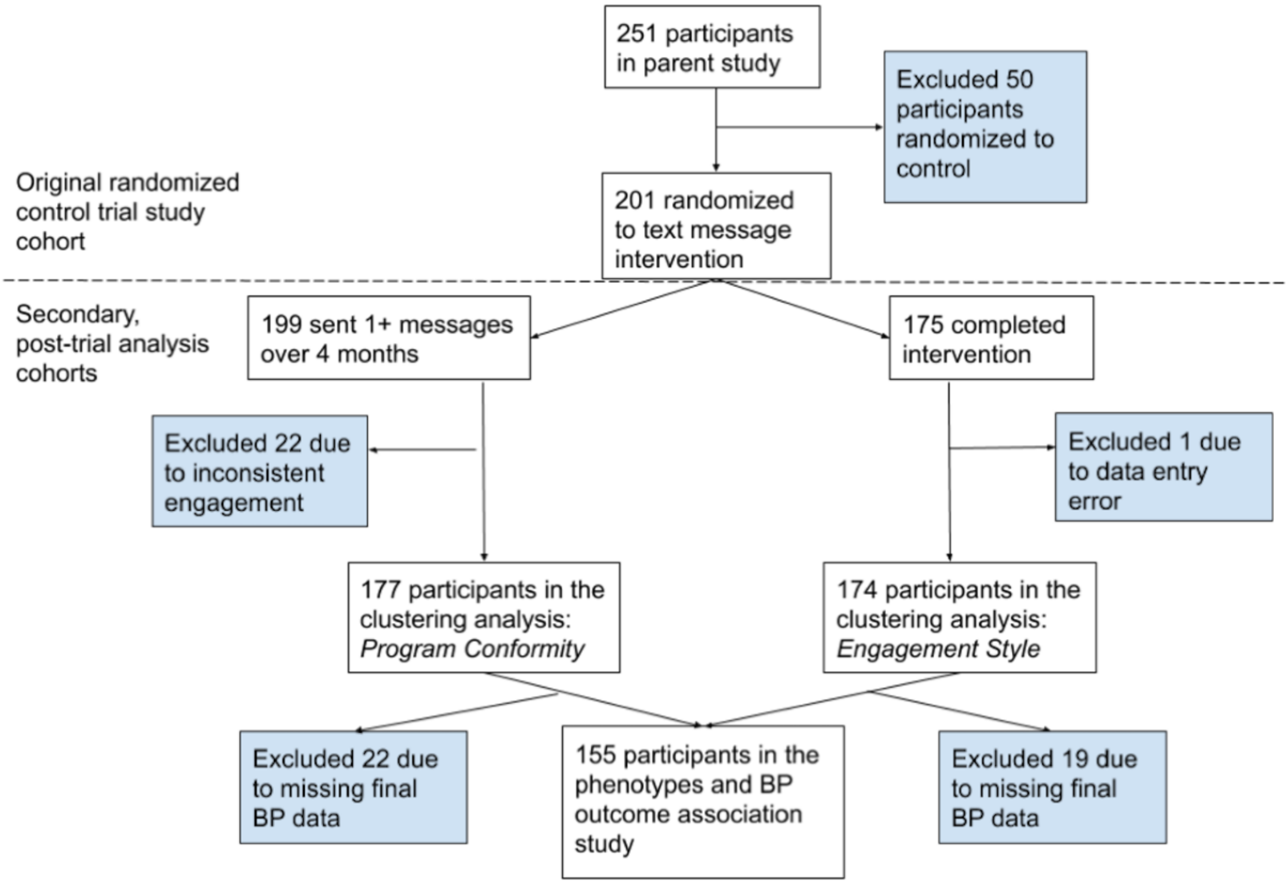
- Proportion of responses to 48 total BP prompts signifying program compliance (BP reporting adherence)
- Average word count per message signifying verbosity (word average), excluding numbers like BP measurements
- Proportion of inbound messages that were not in response to an automated text message prompt (unprompted messages)
- Proportion of patient-submitted messages that triggered an error alert (error rate)

All proportions and word count averages were standardized to a scale between 0 and 1 for the analysis. Again, using the *k*-means clustering algorithm, we identified 3 distinct clusters of engagement styles. To further establish a qualitative understanding of each cluster, we identified the intents—the intended subject of communication (eg, greetings, pleasantry, medication, question, etc)—of all inbound messages from patients that were exceptions to the expected structured text response. Each message was annotated by two members of the research team, applying a common annotation codebook for intents and allowing for multiple intents per message. Discrepancies between the two reviewers were resolved through consensus review. We also described each cluster according to participant age, sex, and race and intervention study arm. Only patients who completed the intervention and had their data entered into the Research Electronic Data Capture database were included in the analysis.

Identifying Associations Between Interaction Phenotypes With Blood Pressure Outcomes

In the original study, final BP was measured at an end-of-study visit. Three BP measurements were taken, and the average of the last 2 measurements represented the final reading. Cutoffs for goal BPs (uncontrolled BP) were in accordance with the American Heart Association/American Stroke Association Eighth Report of the Joint National Committee [20,21]. To identify associations between the interaction phenotypes and target BP, we applied the chi-square test for each cluster and dichotomous BP outcome (controlled vs uncontrolled). Only patients who had all 3 BPs successfully measured and entered into the database were included in the analysis. An overview of participant data subsets from the original study data included for each analysis is shown in Figure 2.

Figure 2. Flow diagram of participants included in each analysis.



Results

Study Population and Data Characteristics

Of the patients who received automated text message reminders, the average age was 50.9 (SD 11.4) years and most were female (142/201, 70.6%; Table 1). From this cohort, we observed 42,263 text message interactions between patients and the automated text messaging platform. Of these interactions, 70.5% (29,791/42,263) of text messages were sent by the automated text messaging platform and 29.5% (12,472/42,263) of text messages were sent by patients. The most messages sent by a single patient was 347 messages. A total of 3.9% (491/12,472) of inbound text messages triggered an error message, and 23.2% (2899/12,472) of all inbound messages were unprompted. A total of 13.9% (1734/12,472) of all messages contained supplemental textual information submitted to the automated text messaging platform.

Table 1. Patient baseline characteristics by intervention group for original randomized control trial sans the control group (n=201).

Characteristics	RM+SS[a] (n=100)	RM[b] (n=101)	*P* value
Gender, female, n (%)	67 (33.3)	75 (37.3)	.07
Age in years, mean (SD)	51.9 (12.5)	50.7 (10.1)	.45
Race, n (%)	—[c]	—	<.001
Black	86 (42.8)	95 (47.3)	—
White	9 (4.5)	3 (1.5)	—
Other	1 (0.5)	5 (2.5)	—
Unknown	4 (2.0)	0 (0)	—
Ethnicity, n (%)	—	—	.96
Hispanic or Latino	0 (0)	1 (0.5)	—
non-Hispanic or Latino	99 (49.3)	100 (49.8)	—
Unknown	1 (0.5)	0 (0)	—
Insurance, n (%)	—	—	<.001
Private	42 (20.9)	56 (27.9)	—
Medicaid	23 (11.4)	28 (13.9)	—
Medicare	33 (16.4)	15 (7.5)	—
None	1 (0.5)	2 (1.0)	—
Unknown	1 (0.5)	0 (0)	—
Texts per patient user, m (SD)	66.9 (27.3)	57.3 (23.3)	.005
Active rate (patient sent at least one message), m (SD) days	139.5 (20.0)	138.3 (15.8)	.54
Processed responses–correctly formatted (BP[d]), mean (SD)	34.4 (11.4)	32.3 (13.0)	.14
Unprocessed responses–error message triggered (BP), mean (SD)	0.8 (1.3)	0.5 (0.7)	.02
Processed messages–medication adherence, mean (SD)	13.0 (4.6)	12.2 (4.9)	.18
Unprocessed messages–medication adherence, mean (SD)	0.9 (1.7)	0.8 (1.3)	.12
Textual patterns for unprocessed messages, mean (SD)			
Character count (per message)	6.7 (10.2)	5.8 (7.0)	.65
Token count (per message)	1.7 (2.7)	1.4 (1.9)	.92
Word count (per message)	1.6 (2.5)	1.3 (1.7)	.75
Number count (per message)	3.7 (3.7)	3.7 (2.1)	.83
Temporal pattern (time of day), mean (SD)			
Morning, per user	50.7 (27.4)	42.3 (24.0)	.02
Afternoon, per user	9.8 (8.1)	8.7 (7.3)	.33
Night, per user	7.2 (7.6)	6.8 (5.6)	.73
Late night, per user	4.1 (3.9)	4.1 (3.6)	.92
Frequency over time, mean (SD)			
Per day	16.9 (11.1)	14.1 (9.9)	<.001
Per week	109.7 (73.6)	87.6 (65.9)	.08
Per month	418.1 (338.6)	361.4 (291.7)	.62
Frequency each month, mean (SD)			
First month, per user	19.7 (7.1)	17.5 (5.5)	.02
Second month, per user	17.7 (7.8)	15.2 (6.3)	.02
Third month, per user	16.5 (6.8)	14.5 (5.5)	.03

Characteristics	RM+SS[a] (n=100)	RM[b] (n=101)	*P* value
Fourth month, per user	14.2 (7.0)	12.6 (5.3)	.08

[a] RM+SS: remote text messaging with social support.

[b] RM: remote text messaging without social support.

[c] Not applicable.

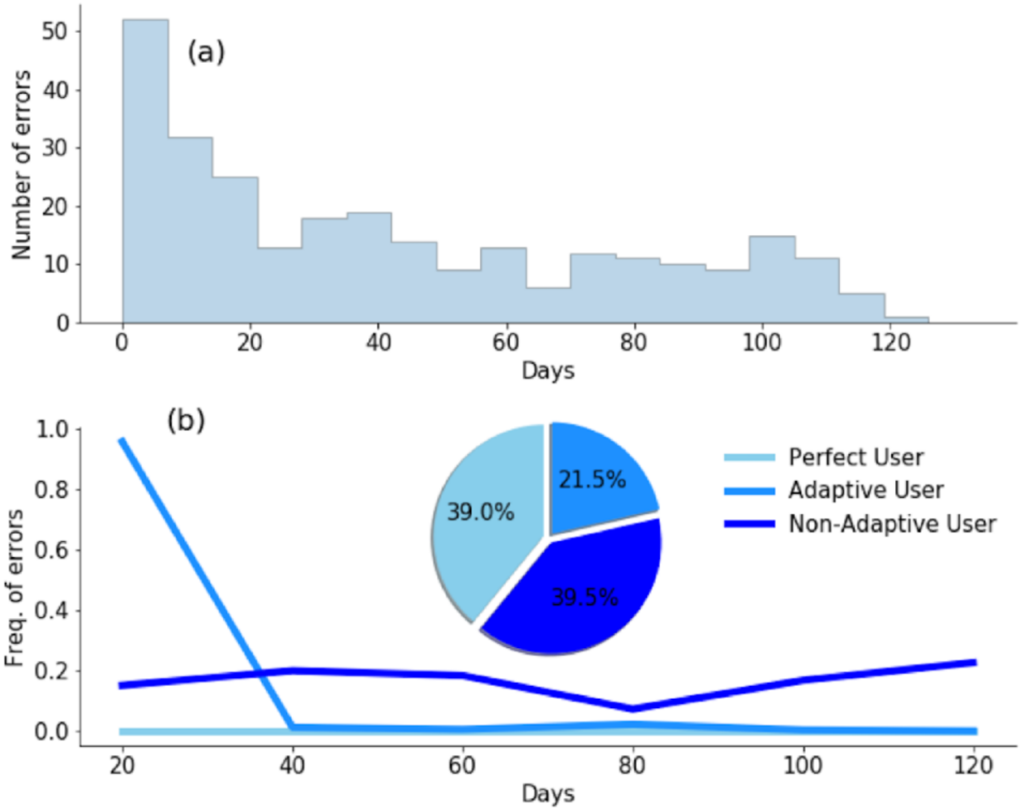
[d] BP: blood pressure.

Characterizing Interaction Phenotypes

Program Conformity

For the 177 patients whose data were analyzed for program conformity, we observed a progressive decline in the number of errors generated by users over the 4 months of their study (Figure 3a). Most errors occurred within the first month overall. Using the *k*-means clustering algorithm, we identified 3 categories of patient program conformity with the text messaging platform defined by the frequency of error messages sent to the patient (Figure 3b). Almost 40% (69/177, 39.0%) of users did not receive any error messages (perfect users); 21.5% (38/177) received error messages within the first month and corrected their submissions for the remainder of the program (adaptive users); and 39.5% (70/177) consistently made errors over the course of all 4 study months (nonadaptive users).

Figure 3. Program conformity clusters: (a) number of unprocessed responses (errors) throughout the 4-month period and (b) frequency of the conformer patterns and trends of errors over time based on program conformity user category.



Engagement Style

For the 174 users whose data were analyzed for engagement style, when we applied the *k*-means clustering method to engagement variables (scaled proportions of messages), we identified 3 notable categories of user engagement styles (Figures 4-6). For each cluster, we present the variables in a radial histogram accompanied by word lists where the size of the intent is associated with its use (Multimedia Appendices 1-3; larger = more frequent use of intent type by users in the cluster). The enthusiast (Figure 4) was characterized by high proportion of BP reporting adherence (range 0.50-1.0), low proportions of errors (range 0.0-0.5), higher proportions of unprompted messages (range 0.25-1.0), and mostly low but dispersed word averages (range 0.0-1.0). Most texts communicated pleasantries ("thank you"), BP with additional text (eg, "today it was," "my bp"), and yes responses (eg, "ok" or "yes, I took it"). Notably absent intents included reports of feeling sick and appointment requests. The student (Figure 5) was characterized by low proportions and high dispersion of BP adherence (range 0.0-0.75), mostly low proportion but dispersed error rates (range 0.0-1.0), low proportion of unprompted messages (range 0.0-0.5), and low proportion but dispersed word averages. Most texts communicated BP, confirmation responses (eg, "yes" or "correct" from users), and communications about their medication(s). Reports about challenging life events were uniquely observed among this cluster (eg, "I'm having financial problems"). The minimalist

(Figure 6) was characterized by higher proportions of BP adherence (range 0.5-1.0), low proportion of errors (range 0.0-0.25), low proportion of unprompted messages (range 0.0-0.25), and low and dispersed word averages. Like the enthusiast, the minimalist texts communicated mostly pleasantries, BP with additional text, and yes responses. Additional unique intents observed include compensation and requests to adjust reminders.

Figure 4. Engagement style: enthusiast—radial distribution of feature values.

Figure 5. Engagement style: student—radial distribution of feature values.

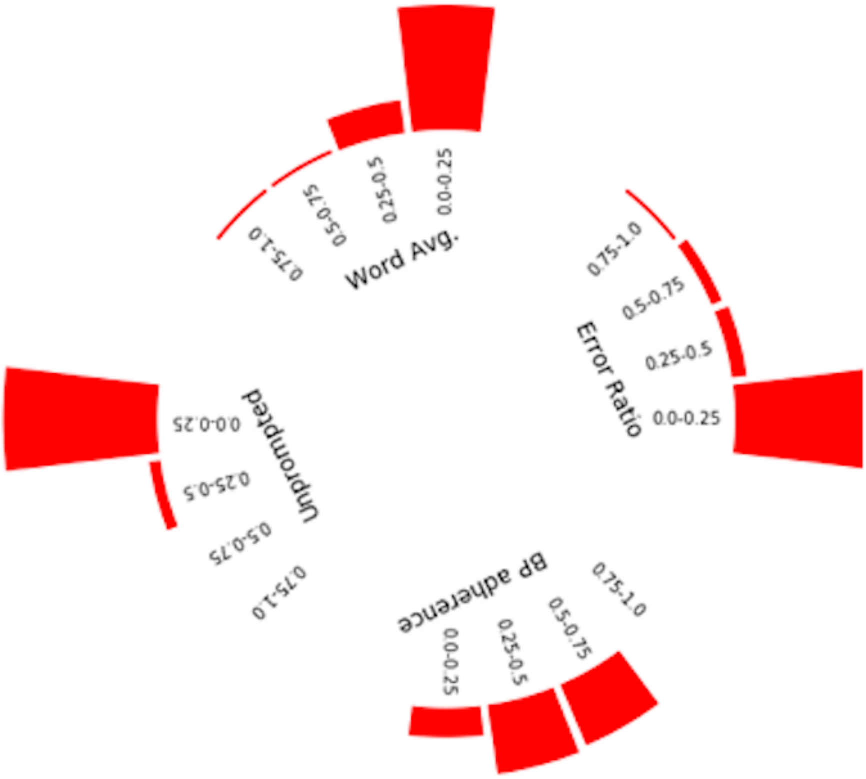


Figure 6. Engagement style: minimalist—radial distribution of feature values.

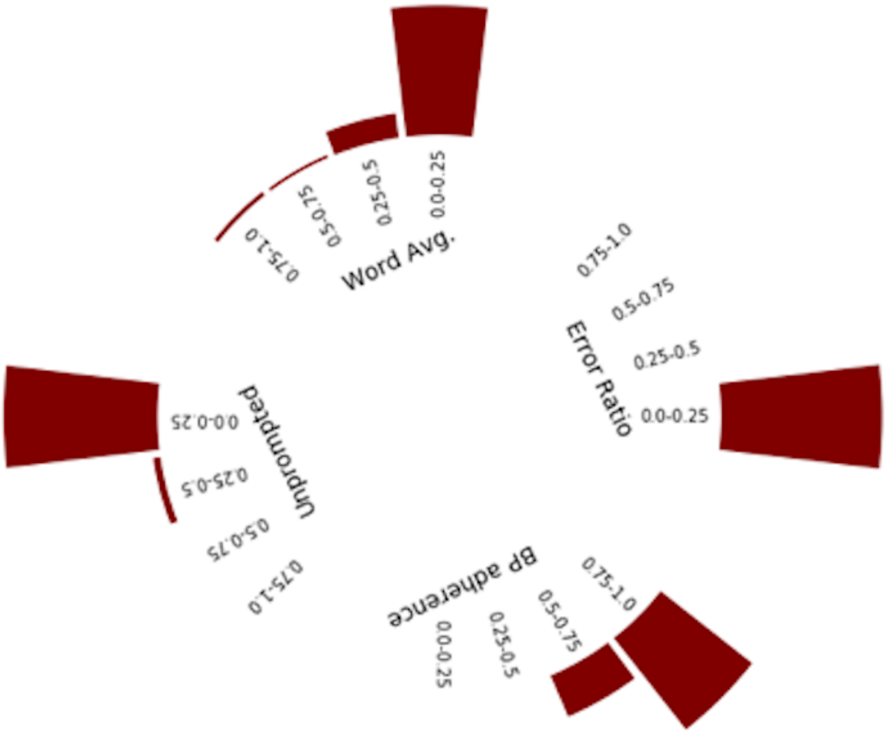


Table 2 further describes engagement style clusters by patient characteristics (race, sex, age) and study arm. For all engagement styles, we observed similar distributions of sex and age. Black was the majority race across all clusters. Most enthusiasts were assigned to the arm with social support. Although most messages conveyed one intent, a notable proportion of messages conveyed two intents across all clusters. More complex messages with multiple intents were infrequent.

Table 2. Distribution of participant characteristics by engagement style.

Characteristic	Enthusiast (n=17)	Student (n=45)	Minimalist (n=112)
Gender, female, n (%)	12 (71)	25 (56)	82 (73)
Age in years, mean (SD)	57.9 (7.3)	47.3 (11.8)	52.4 (11.1)
Race, n (%)			
Black	17 (100)	40 (89)	99 (88)
White	0 (0)	1 (2)	8 (7)
Other	0 (0)	3 (7)	2 (2)
Unknown	0 (0)	1 (2)	3 (3)
Study arm, n (%)			
Remote monitoring	4 (24)	26 (58)	57 (51)
Remote monitoring + social support	13 (77)	19 (42)	55 (49)
Total messages submitted, n (%)			
1 intent (%)	586 (83)	345 (87)	694 (86)
2 intents (%)	114 (16)	43 (11)	103 (13)
3 intents (%)	5 (1)	5 (1)	8 (1)
4+ intents (%)	2 (1)	2 (1)	1 (1)

Identifying Associations Between Interaction Phenotypes With Blood Pressure Outcomes

We characterized the relationship between the 6 phenotypes and BP outcome, controlled versus uncontrolled, for the patients who completed the program and had all 3 end of study visit BPs successfully measured (Table 3). We observed no statistically significant differences among the program conformity clusters. However, among engagement style clusters, a greater proportion of patients in the minimalist cluster achieved controlled BP phenotype (*P*<.001).

Table 3. Association between interaction phenotypes, program conformity, and engagement style, with BP outcomes (n=155).

Interaction phenotype type	Total, n (%)	Users achieving target BP[a], n (%)	*P* value
Program conformity			
Perfect user	51 (32.9)	31 (60.7)	.12
Adaptive user	66 (42.6)	39 (59.1)	.14
Nonadaptive user	38 (24.5)	23 (60.5)	.19
Engagement style			
Enthusiast	17 (11.0)	7 (41.2)	.47
Student	35 (22.6)	19 (54.3)	.61
Minimalist	103 (66.5)	67 (65.0)	<.001

[a] BP: blood pressure.

Discussion

Principal Findings

We identified distinct patterns of patient interaction phenotypes with automated text messages for hypertension management including program conformity and engagement style. We assessed whether these distinct patterns of interaction phenotypes were associated with achieved BP control.

Program Conformity

We first explored heterogeneity in adhering to the structured interactions requested by the automated text messaging platform, using the system's rate of error messages returned to the patient submission as a proxy for program conformity. Perfect users adopted the rules at the beginning of the trial, while adaptive users did so within a month. Nonadaptive users continued to generate error messages throughout the course of the 4-month program.

Our analysis suggests that many patients do not change behavior despite repeated reminders, as the rate of error messages did not change throughout the study's 4-month duration for nonadaptive users. This observation is significant because automated text interventions for hypertension and many other clinical applications predominantly use restricted and structured interactions by limiting patient communication to discrete submissions in a prespecified, standardized format. In our case, there were also time windows for submission so that only a single piece of information (ie, BP or medication adherence) was being collected at a time. This simplified approach makes automated texting technically feasible and approachable across many clinical settings [17,22-24]. Because texting is a largely unrestricted platform, automating a response to correct unstandardized submissions is a way to train users to conform with the program.

However, for even trained users, less restrictive texting or additional means for communication might be desired. A total of 13.9% of messages contained additional text data submitted to the automated texting platform. These messages could have prompted additional engagement between the text messaging platform and the patient or stimulated a conversation between a health care provider and patient. Those implementing automated text systems through structured interactions should bear in mind that a portion of users may never completely adapt to a limited, structured format of response and reply communication. Accounting for heterogeneity in engagement patterns may be one important way to acknowledge the complexity of health-related behaviors [25]. It is likely also the starting point for an equitable approach to automated communication, as those with disabilities or low or limited literacy are more likely to report difficulty texting or following instructions [26]. Automated texting that leads to conversation may result in better data collection, patient engagement, and clinical outcomes.

Engagement Style

Heterogeneity in mobile health engagement styles may explain why such interventions have only demonstrated modest to equivocal clinical impact [27]. In our evaluation, only the minimalist engagement phenotype was associated with significantly better BP control. The patients in this cluster seemed to have a straightforward, business-like relationship with the program, with tight adherence to the original design of the platform. This suggests that for some patients, automated texting with limited structured interaction is sufficient to achieve target BP. The implication is that, while evidence generally suggests text-based interventions should be supplemented by additional care components [28,29], some patients will do just fine with the minimal version. Alternatively, these patients may have achieved target even without any intervention, although all patients were above target at enrollment.

Identifying strategies tailored to other phenotypes may result in better clinical outcomes. The enthusiast demonstrated high levels of engagement, regularly sending BP readings in the appropriate format, as suggested by their low error rates. They also tended to submit more unprompted messages, and compared with other clusters, were hyperverbal, often relaying pleasantries. Despite their high engagement, this phenotype was not associated with significantly improved target BP, although the calculation may have been underpowered. This discrepancy merits further evaluations, and future interventions might consider conversational approaches for this phenotype to improve clinical outcomes.

The student cluster demonstrated more variable BP reporting adherence and error rates, suggesting they faced some challenges in abiding by the rules of the program. Based on their message intents, the students sought more guidance, asked more questions, and had more complex needs. For this group of patients, structured automated programs for BP monitoring may be insufficient to meet all their needs. Importantly, they also relayed clinically significant and meaningful information that was lost and unrecognized by a structured, automated platform. It is possible that the structured format even deterred engagement. In a study of text message communication for mental health among black women, participants commonly cited an impersonal feel and inadequate communication as barriers to adoption [30], and restrictive automated communication may contribute to similar sentiments in this cohort. In comparison with other communication style clusters, the student group tended to skew older and were randomized more frequently without the social support group, which may have affected their behavior. We observed that patients in the RM+SS arm submitted more messages and were more verbose, which could suggest that social support influenced the degree of patient engagement with the platform.

Implications for Next Steps

Our observations have important implications for designing future text messaging and hypertension interventions. Evidence suggests that the most effective interventions have multiple strategies for addressing the informational, behavioral, and social barriers to health and are more effective in larger doses (eg, more time) over a longer period of time [31]. Examples of such strategies include patient education sessions, case management, group support meetings, rewards for meeting BP goals, and pillboxes or medication reminders. However, such complex interventions may also be more resource intensive, and our results suggest that for some, the marginal cost may be unnecessary. Rather than use one universally complex and potentially costly approach, more sustainable interventions might start with automated text support for hypertension management. Additional resources could then be targeted toward those who remain difficult to engage or whose BPs remain uncontrolled; some may require intensive in-person care. Further, our results suggest that identifying patients who send texts outside of the structured format or time windows may present opportunities to engage with patients and identify additional resources to support their care.

Strengths and Limitations

This study has several strengths. To our knowledge, this is the first study to examine variations in the way patients engage with automated text. We leveraged data from a randomized trial and therefore were able to access a large text data file. This analysis included both quantitative and qualitative components, which was possible because we qualitatively coded all text messages.

There are some limitations. This was an exploratory analysis of data from a single center study. No conclusions can be drawn about the specific makeup of clusters in other settings. Clusters defined using another k could identify other phenotype patterns and correlate with BP outcomes, although our sample size limited exploration of greater numbers of k clusters.

Another limitation is that due to incomplete datasets, each of the 3 analyses presented here were conducted in different samples of the study population (n=177, n=174, and n=155, respectively). The sample size was smallest for the analysis of association between text phenotype and BP outcomes because we conducted a complete case analysis. Due to the large number of missing end-of-study BP measurements, imputation of missing values was not justified. Analysis of associations between BP outcomes with behavioral phenotypes was therefore likely limited by small sample size. However, we had a large volume of text messages for our analytic datafile. Also, although the program was designed to be completely automated, on rare occasions it was used to communicate with patients in real time via text regarding study-specific logistics such as study follow-up appointments and reimbursement. This likely altered engagement dynamics for a small number of patients. In addition, the phenotypes described are only pertinent for a short period, as the study was 16 weeks in duration.

Conclusion

Automated texting using a limited, structured interaction format is likely effective in improving BP control for a unique patient phenotype. For others, this format is likely inadequate, and more comprehensive communication and needs assessment may be required. How quickly patients adapt to automation may be less important than how they engage. In particular, patient engagement outside of structured text interactions may signal the need for additional intervention. Future research should identify unique patient phenotypes so that interventions can be tailored accordingly. More research is needed to understand, design, and enhance automated texting platforms so all patients, regardless of phenotype, can reach their BP goals.

Acknowledgments

The project was supported in part by grant number P30AG034546 from the National Institute on Aging. The posttrial analysis presented here was not funded.

Authors' Contributions

SJM and CR provided and helped interpret data. ELA and TD coded text data, and AD analyzed the data. AD and DLM created the data visualization. AD, NSL, CC, KHC, and DM conceptualized the study design. NSL wrote the first draft of the manuscript. All authors revised the manuscript approved the final version.

Conflicts of Interest

KHC is supported in part by grant K12-HS026372-01 from the Agency for Healthcare Research and Quality.

Multimedia Appendix 1
Enthusiast phenotype word cloud.
[PNG File , 105 KB - jmir_v22i12e22493_app1.png]

Multimedia Appendix 2
Minimalist phenotype word cloud.
[PNG File , 141 KB - jmir_v22i12e22493_app2.png]

Multimedia Appendix 3
Student phenotype word cloud.
[PNG File , 83 KB - jmir_v22i12e22493_app3.png]

References

1. Fuchs FD, Whelton PK. High blood pressure and cardiovascular disease. Hypertension 2020 Feb;75(2):285-292. [doi: 10.1161/HYPERTENSIONAHA.119.14240] [Medline: 31865786]
2. Cappuccio FP, Kerry SM, Forbes L, Donald A. Blood pressure control by home monitoring: meta-analysis of randomised trials. BMJ 2004 Jul 17;329(7458):145 [FREE Full text] [doi: 10.1136/bmj.38121.684410.AE] [Medline: 15194600]
3. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]
4. Mobile fact sheet. Washington: Pew Internet and American Life Project; 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-11-03]
5. Langford AT, Solid CA, Scott E, Lad M, Maayan E, Williams SK, et al. Mobile phone ownership, health apps, and tablet use in US adults with a self-reported history of hypertension: cross-sectional study. JMIR Mhealth Uhealth 2019 Jan 14;7(1):e12228 [FREE Full text] [doi: 10.2196/12228] [Medline: 31344667]
6. Hirshberg A, Sammel MD, Srinivas SK. Text message remote monitoring reduced racial disparities in postpartum blood pressure ascertainment. Am J Obstet Gynecol 2019 Sep;221(3):283-285. [doi: 10.1016/j.ajog.2019.05.011] [Medline: 31121137]
7. Lewinski AA, Patel UD, Diamantidis CJ, Oakes M, Baloch K, Crowley MJ, et al. Addressing diabetes and poorly controlled hypertension: pragmatic mhealth self-management intervention. J Med Internet Res 2019 Apr 09;21(4):e12541 [FREE Full text] [doi: 10.2196/12541] [Medline: 30964439]
8. Barsky J, Hunter R, McAllister C, Yeates K, Campbell N, Liu P, et al. Analysis of the implementation, user perspectives, and feedback from a mobile health intervention for individuals living with hypertension (DREAM-GLOBAL): mixed methods study. JMIR Mhealth Uhealth 2019 Dec 09;7(12):e12639 [FREE Full text] [doi: 10.2196/12639] [Medline: 31815678]
9. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
10. Anthony CA, Polgreen LA, Chounramany J, Foster ED, Goerdt CJ, Miller ML, et al. Outpatient blood pressure monitoring using bi-directional text messaging. J Am Soc Hypertens 2015 May;9(5):375-381. [doi: 10.1016/j.jash.2015.01.008] [Medline: 25771023]
11. Irizarry T, Allen M, Suffoletto BP, Einhorn J, Burke LE, Kamarck TW, et al. Development and preliminary feasibility of an automated hypertension self-management system. Am J Med 2018 Sep;131(9):1125-1125 [FREE Full text] [doi: 10.1016/j.amjmed.2018.04.038] [Medline: 29806998]
12. Buis L, Hirzel L, Dawood RM, Dawood KL, Nichols LP, Artinian NT, et al. Text messaging to improve hypertension medication adherence in African Americans from primary care and emergency department settings: results from two randomized feasibility studies. JMIR Mhealth Uhealth 2017 Feb 01;5(2):e9 [FREE Full text] [doi: 10.2196/mhealth.6630] [Medline: 28148474]

13. Mehta SJ, Volpp KG, Troxel AB, Day SC, Lim R, Marcus N, et al. Electronic pill bottles or bidirectional text messaging to improve hypertension medication adherence (Way 2 Text): a randomized clinical trial. J Gen Intern Med 2019 Nov;34(11):2397-2404. [doi: 10.1007/s11606-019-05241-x] [Medline: 31396815]
14. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-Text Adherence Support [StAR]): a single-blind, randomized trial. Circulation 2016 Feb 09;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
15. Meurer WJ, Dome M, Brown D, Delemos D, Oska S, Gorom V, et al. Feasibility of emergency department-initiated, mobile health blood pressure intervention: an exploratory, randomized clinical trial. Acad Emerg Med 2019 May;26(5):517-527. [doi: 10.1111/acem.13691] [Medline: 30659702]
16. Kangovi S, Asch DA. Behavioral phenotyping in health promotion: embracing or avoiding failure. JAMA 2018 Dec 22;319(20):2075-2076. [doi: 10.1001/jama.2018.2921] [Medline: 29710244]
17. Mahmud N, Doshi S, Coniglio M, Clermont M, Bernard D, Reitz C, et al. An automated text message navigation program improves the show rate for outpatient colonoscopy. Health Educ Behav 2019 Dec;46(6):942-946. [doi: 10.1177/1090198119869964] [Medline: 31431077]
18. Hirshberg A, Mahraj K, Srinivas S. Heart safe motherhood: applying innovation methodology for improved maternal outcomes. NEJM Catalyst Mass Med Soc 2019;5(3):1. [doi: 10.1056/CAT.19.0660]
19. Everitt B, Landau S, Leese M, Stahl D. Cluster Analysis. Hoboken: John Wiley & Sons; 2011.
20. Hernandez-Vila E. A review of the JNC 8 blood pressure guideline. Tex Heart Inst J 2015 Jun;42(3):226-228 [FREE Full text] [doi: 10.14503/THIJ-15-5067] [Medline: 26175633]
21. Armstrong C, Joint National Committee. JNC8 guidelines for the management of hypertension in adults. Am Fam Physician 2014 Oct 01;90(7):503-504 [FREE Full text] [Medline: 25369633]
22. Kerbaj J, Toure Y, Soto AA, Boudjema S, Giorgi R, Dufour J, et al. Smartphone text message service to foster hand hygiene compliance in health care workers. Am J Infect Control 2017 Mar 01;45(3):234-239. [doi: 10.1016/j.ajic.2016.10.018] [Medline: 27955945]
23. Uy C, Lopez J, Trinh-Shevrin C, Kwon SC, Sherman SE, Liang PS. Text messaging interventions on cancer screening rates: a systematic review. J Med Internet Res 2017 Aug 24;19(8):e296 [FREE Full text] [doi: 10.2196/jmir.7893] [Medline: 28838885]
24. Agarwal S, Perry HB, Long L, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]
25. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
26. Nouri SS, Avila-Garcia P, Cemballi AG, Sarkar U, Aguilera A, Lyles CR. Assessing mobile phone digital literacy and engagement in user-centered design in a diverse, safety-net population: mixed methods study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e14250 [FREE Full text] [doi: 10.2196/14250] [Medline: 31469083]
27. Gandapur Y, Kianoush S, Kelli HM, Misra S, Urrea B, Blaha MJ, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 01;2(4):237-244 [FREE Full text] [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]
28. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
29. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611. [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
30. McCall T, Schwartz TA, Khairat S. The acceptability of text messaging to help African American women manage anxiety and depression: cross-sectional survey study. JMIR Ment Health 2020 Feb 03;7(2):e15801 [FREE Full text] [doi: 10.2196/15801] [Medline: 31909720]
31. Peacock E, Krousel-Wood M. Adherence to antihypertensive therapy. Med Clin North Am 2017 Jan;101(1):229-245 [FREE Full text] [doi: 10.1016/j.mcna.2016.08.005] [Medline: 27884232]

Abbreviations

BP: blood pressure
RM: remote monitoring only group
RM+SS: remote monitoring + social support group

Edited by G Eysenbach; submitted 17.07.20; peer-reviewed by S Badawy, S Omboni; comments to author 28.08.20; revised version received 12.10.20; accepted 24.10.20; published 03.12.20.

Please cite as:
Davoudi A, Lee NS, Chivers C, Delaney T, Asch EL, Reitz C, Mehta SJ, Chaiyachati KH, Mowery DL
Patient Interaction Phenotypes With an Automated Remote Hypertension Monitoring Program and Their Association With Blood Pressure Control: Observational Study
J Med Internet Res 2020;22(12):e22493
URL: https://www.jmir.org/2020/12/e22493
doi:10.2196/22493
PMID:33270032

Original Paper

Open-Source Technology for Real-Time Continuous Glucose Monitoring in the Neonatal Intensive Care Unit: Case Study in a Neonate With Transient Congenital Hyperinsulinism

Katarina Braune[1,2], MD; Mandy Wäldchen[3], MSc; Klemens Raile[1], MD; Sigrid Hahn[4], MD; Tebbe Ubben[5]; Susanne Römer[4], MD; Daniela Hoeber[6], MD; Nora Johanna Reibel[4], MD; Michael Launspach[2,4], MD; Oliver Blankenstein[7,8], MD; Christoph Bührer[4], MD

[1]Charité – Universitätsmedizin Berlin, Department of Paediatric Endocrinology and Diabetes, Berlin, Germany
[2]Berlin Institute of Health, Berlin, Germany
[3]School of Sociology, University College Dublin, Belfield, Ireland
[4]Charité – Universitätsmedizin Berlin, Department of Neonatology, Berlin, Germany
[5]#dedoc° Diabetes Online Community, Dedoc Labs GmbH, Berlin, Germany
[6]Charité – Universitätsmedizin Berlin, Department of Paediatric Gastroenterology, Nephrology and Metabolic Diseases, Berlin, Germany
[7]Charité – Universitätsmedizin Berlin, Institute for Experimental Paediatric Endocrinology, Berlin, Germany
[8]Charité – Universitätsmedizin Berlin, Newborn Screening Laboratory, Berlin, Germany

Corresponding Author:
Katarina Braune, MD
Charité – Universitätsmedizin Berlin
Department of Paediatric Endocrinology and Diabetes
Augustenburger Platz 1
Berlin, 13353
Germany
Phone: 49 30450616454
Email: katarina.braune@charite.de

Abstract

Background: Use of real-time continuous glucose monitoring (rtCGM) systems has been shown to be a low-pain, safe, and effective method of preventing hypoglycemia and hyperglycemia in people with diabetes of various age groups. Evidence on rtCGM use in infants and in patients with conditions other than diabetes remains limited.

Objective: This case study describes the off-label use of rtCGM and the use of an open-source app for glucose monitoring in a newborn with prolonged hypoglycemia secondary to transient congenital hyperinsulinism during the perinatal period.

Methods: The Dexcom G6 rtCGM system (Dexcom, Inc) was introduced at 39 hours of age. Capillary blood glucose checks were performed regularly. In order to benefit from customizable alert settings and detect hypoglycemic episodes, the open-source rtCGM app xDrip+ was introduced at 9 days of age.

Results: Time in range (45-180 mg/dL) for interstitial glucose remained consistently above 90%, whereas time in hypoglycemia (<45 mg/dL) decreased. Mean glucose was maintained above 70 mg/dL at 72 hours of life and thereafter. Daily sensor glucose profiles showed cyclic fluctuations that were less pronounced over time.

Conclusions: While off-label use of medication is both common practice and a necessity in newborn infants, there are few examples of off-label uses of medical devices, rtCGM being a notable exception. Real-time information allowed us to better understand glycemic patterns and to improve the quality of glycemic control accordingly. Severe hypoglycemia was prevented, and measurement of serum levels of insulin and further lab diagnostics were performed much faster, while the patient's individual burden caused by invasive procedures was reduced. Greater customizability of threshold and alert settings would be beneficial for user groups with glycemic instability other than people with diabetes, and for hospitalized newborn infants in particular. Further research in the field of personal and off-label rtCGM use, efficacy studies evaluating the accuracy of low glucose readings, and studies on the differences between algorithms in translating raw sensor data, as well as customization of commercially available rtCGM systems, is needed.

(J Med Internet Res 2020;22(12):e21770) doi:10.2196/21770

KEYWORDS

open-source; mobile health; continuous glucose monitoring; off-label use; neonatal hypoglycemia; congenital hyperinsulinism; transient hyperinsulinism

Introduction

Transient and persistent congenital hyperinsulinism (CHI) are rare diseases affecting 1 to 2 in 50,000 newborns. At least 12 genetic defects have been identified that result in dysregulated and increased insulin secretion and subsequently severe hypoglycemia, which may be transient or persistent [1,2]. Frequent hypoglycemia is a potentially life-threatening complication of CHI and may lead to permanent brain damage, which may present in developmental delay and mild to severe neurocognitive difficulties such as deficits in attention, memory, and visual and sensorimotor functions in children with CHI [3,4]. Furthermore, prolonged hospitalizations and intense medical regimes pose a constant psychological burden to infants and children with CHI and their caregivers, limiting their participation in social life, kindergarten, and school [2,5]. Therefore, prevention of hypoglycemia is the main treatment goal in patients with CHI [5,6]. Therapeutic regimens are complex and depend on the underlying genotype and phenotype. Treatment options today include regular feeding, pharmacotherapy with diazoxide, somatostatin analogues and glucagon for diffuse forms, and surgery for focal variants of CHI [6].

Real-time continuous glucose monitoring (rtCGM) and intermittently scanned continuous glucose monitoring (isCGM), as designed for and commonly used by people with diabetes, provide information on current interstitial glucose levels and trends, and alert the user of current or predicted hypoglycemia and hyperglycemia as well as rapid changes in glucose levels. With regular measurements up to every 5 minutes, they also provide statistics on time spent in hypoglycemia, in hyperglycemia, and within range.

Evidence on rtCGM use for glycemic management in neonates is limited, and there is no evidence of calibration-free rtCGM systems in infants so far. Benefits with respect to glycemic control [7] and less procedural pain [8] have been previously described for very low-birth-weight preterm infants in a feasibility study. Furthermore, a study found continuous tissue glucose monitoring to be helpful to identify infants at risk for metabolic instabilities [9]. However, knowledge remains limited for infants and other patient groups at risk of hypoglycemia so far [10,11].

The off-label use of rtCGM in children with CHI was first described in 2004 [12]. Further studies describe rtCGM and isCGM as safe, effective, convenient, and less painful methods of glucose monitoring in CHI patients that provide helpful information for therapeutic decisions [13,14]. Caregivers found the indication of glucose trends to be helpful, felt more confident in everyday management of the condition, and felt less worried about the occurrence of asymptomatic hypoglycemia [12,14]. However, older rtCGM/isCGM systems were reported to have limited accuracy, especially for hypoglycemic levels, and therefore not recommended to completely replace capillary blood glucose (CBG) testing, but rather seen as an adjunctive, beneficial therapy [13,15].

To our knowledge, this case study describes the first use of the rtCGM app xDrip+ in a hospital and intensive care unit (ICU) setting and the first off-label use of rtCGM in a newborn with severe hypoglycemia secondary to transient CHI soon after birth.

Methods

Clinical Case

A male infant was delivered full term with a birth weight of 3575 g. An emergency cesarean delivery was performed in response to fetal heart rate monitoring indicating fetal distress. The infant presented with bradypnea, floppiness, and sporadic myoclonic jerks and was transiently treated with noninvasive positive pressure resuscitation. A CBG level of 12 mg/dL was detected at 60 minutes of life. Despite feeding attempts and buccal administration of 40% dextrose gel, the blood glucose level decreased further to 1 mg/dL at 120 minutes of life. The patient was started on continuous intravenous glucose and admitted to the neonatal intensive care unit (NICU) for further diagnostics and treatment. The continuous intravenous glucose administration required to maintain blood glucose levels above 50 mg/dL was incrementally increased from 8 mg/kg/min to over 18 mg/kg/min. This required placement of a central venous line at 37 hours of age to allow for administration of hypertonic glucose solutions. In addition, the infant received copious enteral feedings.

The child’s father and the paternal uncle were both reported to have experienced transient postnatal hypoglycemia, requiring hospitalization for 3 weeks after birth. The child’s father reported having often experienced mild hypoglycemia symptoms such as hunger and craving for sweets when fasting. However, blood glucose levels were reported to be within normal range during routine checkups as an adult. There was no history of diabetes in either of the parents, and screening for gestational diabetes during pregnancy was reported to be negative. However, the paternal grandmother was reported to have been diagnosed with type 1 diabetes at 7 years of age.

Genetic Testing

Next-generation sequencing (quality level type A; SureSelect XT, Custom Constitutional Panel 17 Mb; Agilent Technologies, Inc) of a CHI panel, including *ABCC8* (GenBank NM_000352.4), *HADH* (GenBank NM_005327.4), *HNF4A* (GenBank NM_175914.4), *KCNJ11* (GenBank NM_000525.3), *KMT2D* (GenBank NM_003482), and *UCP2* (GenBank NM_003355.2), was performed by Labor Berlin, Germany.

Real-Time Glucose Monitoring

To reduce the frequency of capillary blood samples and estimate the right moment for further lab diagnostics, the Dexcom G6 rtCGM system (Dexcom, Inc) was introduced for off-label use at 39 hours of age. In addition, CBG checks were performed regularly, at least 3 times per day (Multimedia Appendix 1). All decisions concerning the therapeutic regimen were based on confirmatory CBG tests. Sensors were placed at the lateral sides of thighs or upper arms (Figures 1 and 2) where enough fat and muscle tissue were present. Sensors were replaced every 7 days.

During the first week, the sensor was paired with a Dexcom G6 handheld device with customizable alert settings; however, the urgent low glucose alarm threshold at a glucose level of 55 mg/dL, which was frequently passed in this neonate but does not meet the requirements in neonates within the first week of life, could not be further reduced. Furthermore, glucose levels below 40 mg/dL could not be further differentiated. In order to benefit from customizable alert settings and include potential hypoglycemic levels below 40 mg/dL, the informed decision was made to pair the sensors with the caregivers' personal Android phone using the app xDrip+.

Figure 1. Upper thigh as a suitable rtCGM application site in neonates.

Figure 2. Upper arm as a suitable rtCGM application site in neonates.

Specifics of the xDrip+ Algorithm

xDrip+ [16] is an open-source product with no regulatory approval that allows the user to choose between the approved native algorithm, which runs on the Dexcom G6 transmitter for translating uncalibrated transmitter data into interstitial glucose level estimations, and its own algorithm, which approximates a linear function using the method of weighted least squares regression by weighting calibration points based on their value, sensor and calibration age, the current glucose trend, a variability heuristic, and other factors. Furthermore, low glucose alarm settings are fully customizable. In this case study, the xDrip+ algorithm was used with a minor modification that allowed sensor glucose levels below 40 mg/dL to be received, although accuracy was expected to be limited in this range. Alerts were set at 45 mg/dL for low sensor glucose levels and at 140 mg/dL for high sensor glucose levels.

Data Analysis

For the first 7 days of sensor data, rtCGM readings were exported from the Dexcom handheld device using the Dexcom Clarity software. For the following days when xDrip+ was used, sensor data was exported through the built-in data export function in the xDrip+ app. For the analysis of sensor data, we computed statistical properties (arithmetic mean and standard deviation) in the programming language Python 3.7. Furthermore, time in hypoglycemia and time in range were calculated as follows:

As there is currently no standardized definition for treatment consequences in response to low and high glucose levels in neonates, but their physiological range tends to be lower than in older individuals [9], sensor glucose in range was defined by a lower threshold of 45 mg/dL (2.5 mmol/L) and an upper threshold of 180 mg/dL (10 mmol/L). Therefore, time in hypoglycemia was calculated based on sensor readings under 45 mg/dL (2.5 mmol/L).

Results

Diagnostic Workup

The infant responded well to a single subcutaneous injection of 30 µg/kg of glucagon (GlucaGen Hypokit; Novo Nordisk A/S). Therefore, a glycogen storage disease was unlikely to be the underlying condition and congenital hyperinsulinism was suspected. According to CHI treatment guidelines [6], a continuous subcutaneous glucagon infusion was initiated at 64 hours of age and gradually increased (up to 27.2 µg/kg/h) (Multimedia Appendix 1). Glucagon dosage was adjusted (increased or lowered) according to rtCGM glucose using defined thresholds for each action; thus, the continuous glucose infusion could be gradually reduced and completely terminated by 13 days of age. In the following days, glucagon administration was gradually decreased and completely terminated 17 days of age. That treatment algorithm successfully prevented occurrence of any symptomatic hypoglycemia (eg, seizures, hypothermia, hypotonia, bradypnea). No other drugs such as diazoxide or lanreotide were administered.

Initial hypoglycemia screening revealed hyperinsulinemia with no ketonemia present (blood glucose 29 mg/dL, insulin 9.03 mU/L, C-peptide 2.34 µg/L). After terminating the intravenous glucose and the subcutaneous glucagon infusion and with increasing improvement of oral feeding and hence regular and reliable enteral glucose uptake, no new hypoglycemic episodes occurred between days 17 and 20. Consequently, we suspected normalization of insulin secretion, which was substantiated by an analysis after a strict fasting period of 6.5 hours (blood glucose 86 mg/dL, insulin 1.5 mU/L, C peptide 0.54 µg/L, ketone bodies 0.4 mmol/L) on day 20. With these results, we were able to terminate rtCGM use and discharge the infant on the same day in good health.

Genetic testing did not reveal any disease-relevant pathologic variation in genes that are most frequently associated with CHI: *ABCC8* (GenBank NM_000352.4), *HADH* (GenBank NM_005327.4), HNF4A (GenBank NM_175914.4), *KCNJ11* (GenBank NM_000525.3), and *UCP2* (NM_003355.2). However, we detected a novel, highly conserved, and heterozygous missense variant in the *KMT2D* gene (c.3976C>T, R1326W), which was interpreted as a pathogenic variant within this analysis and might be causative in this particular case of a Kabuki syndrome with predominantly transient, congenital hyperinsulinism [17,18].

Glycemic Outcomes

Time in range (sensor glucose readings between 45 and 180 mg/dL or 2.5 and 10 mmol/L) remained consistent above 90%, whereas time in hypoglycemia (<45 mg/dL or <2.5 mmol/L) decreased, and mean glucose was maintained above 70 mg/dL or 3.8 mmol/L at 72 hours of life and after (Table 1). Daily sensor glucose profiles (Multimedia Appendix 1) show cyclic fluctuations exceeding 45 mg/dL (2.5 mmol/L) on most days and occasionally 180 mg/dL (10 mmol/L). These cyclic fluctuations were found to be less pronounced over time.

rtCGM readings were found to be mostly accurate compared to CBG (Multimedia Appendix 1). Significant deviations (>15%) occurred within the first 24 hours after a new sensor was placed. After recalibration of the rtCGM, accuracy of rtCGM readings was satisfactory (<15% deviations).

Table 1. Analysis of rtCGM readings within the first 16 days of life in a neonate with transient CHI, including mean sensor glucose (SD), time in hypoglycemia (defined as the percentage of sensor readings below 45 mg/dL), and time in range (defined as the percentage of sensor readings between 45 mg/dL and 180 mg/dL).

Patient age, d (h)	Sensor readings, n	Glucose, mg/dL, mean (SD)	Readings in range, n	TIR[a], %	Readings in hypoglycemia, n	TIH[b], %
2 (39-48)	74	68 (16)	71	95.9	3	4.1
3 (49-72)	291	69 (29)	240	82.5	51	17.5
4 (73-96)	288	98 (34)	288	100.0	0	0.0
5 (97-120)	290	79 (16)	286	98.6	4	1.4
6 (121-144)	291	82 (17)	280	96.2	11	3.8
7 (145-168)	282	94 (36)	274	97.2	8	2.8
8 (169-192)	287	77 (21)	270	94.1	17	5.9
9 (193-216)	203	85 (29)	190	93.6	13	6.4
10 (217-240)	213	81 (22)	197	92.5	16	7.5
11 (241-264)	284	81 (25)	261	91.9	23	8.1
12 (265-288)	281	86 (24)	274	97.5	7	2.5
13 (289-312)	296	81 (19)	284	95.9	12	4.1
14 (313-336)	288	73 (20)	265	92.0	23	8.0
15 (337-360)	248	72 (18)	235	94.8	13	5.2
16 (361-372)	138	83 (17)	138	100.0	0	0.0

[a] TIR: time in range.

[b] TIH: time in hypoglycemia.

Discussion

Principal Findings

This case study describes the use of a Dexcom G6 rtCGM system in a neonate with transient CHI during the perinatal period, using two different algorithms. The real-time information provided by the rtCGM system allowed us to better understand glycemic patterns of the underlying condition and to improve the quality of glycemic control of our patient accordingly. Severe hypoglycemia was successfully prevented and measurement of serum levels of insulin and further lab diagnostics were performed much faster, while the patient's individual burden caused by invasive procedures could be reduced.

To our knowledge, this is the first report on the use of open-source apps for glucose monitoring such as xDrip+ in a hospital and ICU setting, as well as the first study reporting an off-label use of a precalibrated rtCGM system in a neonate. Furthermore, we described the first use of rtCGM in an infant with transient CHI in the most critical phase during the first days of life.

While off-label use of medication is both common practice and a necessity in newborn infants, there are few examples of off-label uses of medical devices, rtCGM being a notable exception [7,8,19-24]. CBG tests are the current standard of care in the management of neonatal hypoglycemia. However, CBG has notable limitations. Heel pricks, associated with pain, are required every time a blood sample is taken. Furthermore, it provides only single point measurements without a continuous profile on glucose trends, dynamics, and interactions with drugs and meals. Conversely, rtCGM records any hypoglycemic episode in a quantitative fashion, displays detailed pharmacodynamics of single drugs, and consequently can help to determine a drug's long-term requirement. We assume that rtCGM use could be extremely helpful in any hospital or home setting and in this case, could help reduce the child's risk for neurological long-term complications [3,4,25]. Our experiences with the presented case indicate that greater customizability of threshold and alert settings would be beneficial for user groups with glycemic instability other than people with diabetes, and for hospitalized infants in particular. Further potential use cases may include children of mothers with gestational diabetes, children with metabolic disorders, children with neonatal diabetes mellitus, preterm infants, and infants with difficulties in enteral feeding.

Strengths and Limitations

It is important to note that although rtCGM provides information about glucose dynamics and variability, these systems might be less accurate when it comes to direct comparison with point-of-care testing of blood glucose levels. In order to keep invasive diagnostics for the patient reduced to a minimum, there were not enough CBG data points available to allow for a robust calculation of the mean absolute relative difference. This limits our ability to evaluate the accuracy of rtCGM readings, particularly of glucose levels <40 mg/dL. As rtCGM systems correlate the electric signal of the sensor with interstitial glucose concentrations in a linear fashion, they are not designed to detect the low glucose levels that occur in newborn infants after birth. Moreover, alert settings are not fully customizable to threshold values that seem reasonable to use in infants, who have a much lower physiological glucose range compared to older children and adults [9].

The app xDrip+, which was used in this case study, originates from an online community of people living with type 1 diabetes. The Android application package needed for setup, the app's source code, and instructions on how to install and use the app are openly available on the internet. Recently, open-source rtCGM apps such as xDrip+ for Android phones [16] and Spike for Apple iOS [26] have become increasingly popular and are used by thousands of people in the worldwide type 1 diabetes community [27], especially by those living in countries where commercial rtCGM systems are not accessible or not reimbursed by the healthcare system [28]. Despite their popularity, so far neither of the aforementioned open-source rtCGM apps have been tested in clinical trials or been approved by a regulatory body. The creators of these apps and the online community around them publicly disclose that they should be used with caution at all times and at one's own risk. The controlled environment of a NICU was an ideal setting to observe outcomes and potential benefits of the off-label use while ensuring minimal risk for the patient at the same time. We found that the app was also a more suitable solution that best met the child's medical needs as well as healthcare professionals' and the caregivers' expectations on relevant alarms in case of glucose excursions. However, the safety and accuracy of xDrip+ as such has not been systematically investigated so far, and its use has only been described in conjunction with observational and self-reported data from open-source automated insulin delivery system users [29,30]. Further research on personal and off-label rtCGM use and differences between native and alternative algorithms in translating raw sensor data is needed, as well as customization of commercially available rtCGM systems.

Conclusions

The use of rtCGM can be considered in neonatal patients at risk of hypoglycemia, such as infants with transient or permanent CHI, to reduce the frequency of blood glucose measurements and focus them on potential decision-making points. However, these devices have been designed to be used by people with diabetes and are currently not approved for use in children younger than 2 years of age. Other use cases, such as infants with glycemic instability and other patients with rare conditions that might benefit from continuous glucose monitoring, are currently not being addressed sufficiently by the medical device market. Further investigation on the use and accuracy of rtCGM in wider patient groups, as well as further customization of rtCGM systems, is needed to address unmet needs of wider population groups.

Acknowledgments

We would like to thank the developers of xDrip+ of the diabetes online community for their cooperation, helpful comments, and thorough review of the paper.

The caregivers gave their consent for the off-label use of xDrip+ and for the publication of this case study.

The authors received no financial funding for the research conducted. We acknowledge support from the German Research Foundation and the Open Access Publication Funds of Charité – Universitätsmedizin Berlin for the publication of this paper.

Authors' Contributions

KB, MW, and SH performed the literature search. KB, OB, and CB designed the study design and methods. KB, ML, DH, NJR, and SH collected the data. KB and MW performed the data analysis. KB wrote the initial draft for the paper. All authors have reviewed and revised the manuscript and have read and approved the final version of the paper.

Conflicts of Interest

All authors have completed the Unified Competing Interest form and declare the following: KB received research grants from the Berlin Institute of Health Junior Clinician Scientist and Digital Clinician Scientist program, the European Commission's Horizon 2020 Research and Innovation program, the Wellcome Trust, Stiftung Charité, and the German Diabetes Association. She has served as a speaker and advisory board member for Medtronic Diabetes and Hi.Health and received fees for medical consulting and public speaking from Roche Diabetes Care, Dexcom, Medtronic Diabetes, Diabeloop, BCG Digital Ventures, and Bertelsmann Stiftung; all outside the submitted work. MW reports research grants from the European Commission's Horizon 2020 Research and Innovation program outside the submitted work and was employed by Medtronic Diabetes Germany until 2018. TU reports research grants from the European Commission's Horizon 2020 Research and Innovation program outside the submitted work. ML received research grants from the Berlin Institute of Health and the Berlin Cancer Society outside the submitted work. CB is supported by the German Federal Department of Research and Education and has received lecture fees from Chiesi paid to his employer, both outside the submitted work. All other coauthors have no conflict of interest to declare.

Multimedia Appendix 1

Daily profiles of sensor glucose and therapeutic management of a newborn with transient CHI. Sensor glucose readings [mg/dL] are shown in blue, i.v. glucose distribution in green [mg/kg/min] and s.c. glucagon infusion in orange [μg/kg/h]. Low sensor glucose threshold (45 mg/dL) is shown in red. Capillary blood glucose (CBG) measurements are marked in magenta.
[DOCX File , 899 KB - jmir_v22i12e21770_app1.docx]

References

1. Demirbilek H, Hussain K. Congenital Hyperinsulinism: Diagnosis and Treatment Update. J Clin Res Pediatr Endocrinol 2017 Dec 30;9(Suppl 2):69-87 [FREE Full text] [doi: 10.4274/jcrpe.2017.S007] [Medline: 29280746]
2. Arnoux J, de Lonlay P, Ribeiro M, Hussain K, Blankenstein O, Mohnike K, et al. Congenital hyperinsulinism. Early Hum Dev 2010 May;86(5):287-294. [doi: 10.1016/j.earlhumdev.2010.05.003] [Medline: 20550977]
3. Ludwig A, Enke S, Heindorf J, Empting S, Meissner T, Mohnike K. Formal Neurocognitive Testing in 60 Patients with Congenital Hyperinsulinism. Horm Res Paediatr 2018;89(1):1-6. [doi: 10.1159/000481774] [Medline: 29151084]
4. Muukkonen L, Männistö J, Jääskeläinen J, Hannonen R, Huopio H. The effect of hypoglycaemia on neurocognitive outcome in children and adolescents with transient or persistent congenital hyperinsulinism. Dev Med Child Neurol 2019 Apr;61(4):451-457 [FREE Full text] [doi: 10.1111/dmcn.14039] [Medline: 30246438]
5. Minakova E, Chu A. Congenital Hyperinsulinism. Pediatr Ann 2017 Nov 01;46(11):e409-e414. [doi: 10.3928/19382359-20171020-01] [Medline: 29131920]
6. AWMF. Diagnostik und Therapie des Kongenitalen Hyperinsulinismus (KHI). AWMF Leitlinien. URL: https://www.awmf.org/uploads/tx_szleitlinien/174-012l_S1_Diagnostik-Therapie-Kongenitaler-Hyperinsulinismus-KHI_2020-01.pdf [accessed 2020-06-20]
7. Beardsall K, Thomson L, Elleri D, Dunger DB, Hovorka R. Feasibility of automated insulin delivery guided by continuous glucose monitoring in preterm infants. Arch Dis Child Fetal Neonatal Ed 2020 May;105(3):279-284. [doi: 10.1136/archdischild-2019-316871] [Medline: 31399480]
8. Galderisi A, Lago P, Steil GM, Ghirardo M, Cobelli C, Baraldi E, et al. Procedural Pain during Insertion of a Continuous Glucose Monitoring Device in Preterm Infants. J Pediatr 2018 Sep;200:261-264.e1. [doi: 10.1016/j.jpeds.2018.03.040] [Medline: 29861315]
9. Mola-Schenzle E, Staffler A, Klemme M, Pellegrini F, Molinaro G, Parhofer KG, et al. Clinically stable very low birthweight infants are at risk for recurrent tissue glucose fluctuations even after fully established enteral nutrition. Arch Dis Child Fetal Neonatal Ed 2015 Mar;100(2):F126-F131. [doi: 10.1136/archdischild-2014-306168] [Medline: 25381093]
10. Ogilvy-Stuart AL, Beardsall K. Management of hyperglycaemia in the preterm infant. Arch Dis Child Fetal Neonatal Ed 2010 Mar;95(2):F126-F131. [doi: 10.1136/adc.2008.154716] [Medline: 20231218]
11. McKinlay CJD, Chase JG, Dickson J, Harris DL, Alsweiler JM, Harding JE. Continuous glucose monitoring in neonates: a review. Matern Health Neonatol Perinatol 2017;3:18 [FREE Full text] [doi: 10.1186/s40748-017-0055-z] [Medline: 29051825]
12. Conrad SC, Mastrototaro JJ, Gitelman SE. The use of a continuous glucose monitoring system in hypoglycemic disorders. J Pediatr Endocrinol Metab 2004 Mar;17(3):281-288. [doi: 10.1515/jpem.2004.17.3.281] [Medline: 15112904]
13. Rayannavar A, Elci OU, Mitteer L, De León DD. Continuous Glucose Monitoring Systems: Are They Useful for Evaluating Glycemic Control in Children with Hyperinsulinism? Horm Res Paediatr 2019;92(5):319-327. [doi: 10.1159/000506230] [Medline: 32208390]
14. Saif M, Kapoor A, Kochar IPS, Jindal R. Continuous glucose monitoring system for congenital hyperinsulinemia. Indian Pediatr 2013 Apr;50(4):421-422 [FREE Full text] [doi: 10.1007/s13312-013-0103-3] [Medline: 23665604]
15. Alsaffar H, Turner L, Yung Z, Didi M, Senniappan S. Continuous Flash Glucose Monitoring in children with Congenital Hyperinsulinism; first report on accuracy and patient experience. Int J Pediatr Endocrinol 2018;2018:3 [FREE Full text] [doi: 10.1186/s13633-018-0057-2] [Medline: 29599801]
16. Github - xDrip Plus. URL: https://github.com/NightscoutFoundation/xDrip [accessed 2020-06-20]
17. Yap KL, Johnson AEK, Fischer D, Kandikatla P, Deml J, Nelakuditi V, et al. Congenital hyperinsulinism as the presenting feature of Kabuki syndrome: clinical and molecular characterization of 9 affected individuals. Genet Med 2019 Jan;21(1):233-242. [doi: 10.1038/s41436-018-0013-9] [Medline: 29907798]
18. Hoermann H, El-Rifai O, Schebek M, Lodefalk M, Brusgaard K, Bachmann N, et al. Comparative meta-analysis of Kabuki syndrome with and without hyperinsulinaemic hypoglycaemia. Clin Endocrinol (Oxf) 2020 Jun 13:346-354. [doi: 10.1111/cen.14267] [Medline: 32533869]
19. Nally LM, Bondy N, Doiev J, Buckingham BA, Wilson DM. A Feasibility Study to Detect Neonatal Hypoglycemia in Infants of Diabetic Mothers Using Real-Time Continuous Glucose Monitoring. Diabetes Technol Ther 2019 Apr;21(4):170-176. [doi: 10.1089/dia.2018.0337] [Medline: 30839229]
20. Pinchefsky EF, Hahn CD, Kamino D, Chau V, Brant R, Moore AM, et al. Hyperglycemia and Glucose Variability Are Associated with Worse Brain Function and Seizures in Neonatal Encephalopathy: A Prospective Cohort Study. J Pediatr 2019 Jun;209:23-32. [doi: 10.1016/j.jpeds.2019.02.027] [Medline: 30982528]
21. Tomotaki S, Toyoshima K, Shimokaze T, Kawai M. Reliability of real-time continuous glucose monitoring in infants. Pediatr Int 2019 Oct;61(10):1001-1006. [doi: 10.1111/ped.13961] [Medline: 31287607]
22. Cumberpatch AR, Weston PJ, Harding JE, Harris DL. Parents of babies who participated in an invasive clinical study report a positive experience: the Glucose in Well Babies (GLOW) study. Arch Dis Child Fetal Neonatal Ed 2020 Jan;105(1):4-7. [doi: 10.1136/archdischild-2019-317417] [Medline: 31666312]
23. Harris DL, Weston PJ, Gamble GD, Harding JE. Glucose Profiles in Healthy Term Infants in the First 5 Days: The Glucose in Well Babies (GLOW) Study. J Pediatr 2020 May 04:34-41. [doi: 10.1016/j.jpeds.2020.02.079] [Medline: 32381469]

24. Saw H, Yao N, Chiu C, Chen J. The value of real-time continuous glucose monitoring in premature infants of diabetic mothers. PLoS One 2017;12(10):e0186486 [FREE Full text] [doi: 10.1371/journal.pone.0186486] [Medline: 29036213]
25. McKinlay CJD, Alsweiler JM, Ansell JM, Anstice NS, Chase JG, Gamble GD, CHYLD Study Group. Neonatal Glycemia and Neurodevelopmental Outcomes at 2 Years. N Engl J Med 2015 Oct 15;373(16):1507-1518 [FREE Full text] [doi: 10.1056/NEJMoa1504909] [Medline: 26465984]
26. Spike: Get the most out of your CGM. Spike. URL: https://github.com/SpikeApp/Spike [accessed 2020-06-20]
27. Lee JM, Newman MW, Gebremariam A, Choi P, Lewis D, Nordgren W, et al. Real-World Use and Self-Reported Health Outcomes of a Patient-Designed Do-it-Yourself Mobile Technology System for Diabetes: Lessons for Mobile Health. Diabetes Technol Ther 2017 Apr;19(4):209-219. [doi: 10.1089/dia.2016.0312] [Medline: 28245152]
28. Pfiester E. Cost and Rationing of Insulin and Diabetes Supplies: Findings from the T1International Patient Survey. T1International. URL: https://www.t1international.com/media/assets/file/T1International_Report_-_Costs_and_Rationing_of_Insulin__Diabetes_Supplies_2.pdf [accessed 2020-06-23]
29. Petruzelkova L, Soupal J, Plasova V, Jiranova P, Neuman V, Plachy L, et al. Excellent Glycemic Control Maintained by Open-Source Hybrid Closed-Loop AndroidAPS During and After Sustained Physical Activity. Diabetes Technol Ther 2018 Dec;20(11):744-750. [doi: 10.1089/dia.2018.0214] [Medline: 30285476]
30. Braune K, O'Donnell S, Cleal B, Lewis D, Tappe A, Willaing I, et al. Real-World Use of Do-It-Yourself Artificial Pancreas Systems in Children and Adolescents With Type 1 Diabetes: Online Survey and Analysis of Self-Reported Clinical Outcomes. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e14087 [FREE Full text] [doi: 10.2196/14087] [Medline: 31364599]

Abbreviations

CBG: capillary blood glucose
CHI: congenital hyperinsulinism
ICU: intensive care unit
isCGM: intermittently scanned continuous glucose monitoring
NICU: neonatal intensive care unit
rtCGM: real-time continuous glucose monitoring

Edited by G Eysenbach; submitted 12.07.20; peer-reviewed by K Gajewska, M Clements; comments to author 16.08.20; revised version received 15.09.20; accepted 26.10.20; published 04.12.20.

Please cite as:
Braune K, Wäldchen M, Raile K, Hahn S, Ubben T, Römer S, Hoeber D, Reibel NJ, Launspach M, Blankenstein O, Bührer C
Open-Source Technology for Real-Time Continuous Glucose Monitoring in the Neonatal Intensive Care Unit: Case Study in a Neonate With Transient Congenital Hyperinsulinism
J Med Internet Res 2020;22(12):e21770
URL: http://www.jmir.org/2020/12/e21770/
doi:10.2196/21770
PMID:33275114

Original Paper

Using an Ontology to Facilitate More Accurate Coding of Social Prescriptions Addressing Social Determinants of Health: Feasibility Study

Anant Jani[1], PhD; Harshana Liyanage[2], PhD; Cecilia Okusi[2], MRES; Julian Sherlock[2], BSc; Uy Hoang[2], MBBS; Filipa Ferreira[2], PhD; Ivelina Yonova[2], MA; Simon de Lusignan[2], MBBS

[1] Oxford Martin School, University of Oxford, Oxford, United Kingdom

[2] Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Simon de Lusignan, MBBS
Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building, Radcliffe Observatory Quarter
Woodstock Rd
Oxford, OX2 6GG
United Kingdom
Phone: 44 01865 617 283 ext 17
Email: simon.delusignan@phc.ox.ac.uk

Abstract

Background: National Health Service (NHS) England supports social prescribing in order to address social determinants of health, which account for approximately 80% of all health outcomes. Nevertheless, data on ongoing social prescribing activities are lacking. Although NHS England has attempted to overcome this problem by recommending 3 standardized primary care codes, these codes do not capture the social prescribing activity to a level of granularity that would allow for fair attribution of outcomes to social prescribing.

Objective: In this study, we explored whether an alternative approach to coding social prescribing activity, specifically through a social prescribing ontology, can be used to capture the social prescriptions used in primary care in greater detail.

Methods: The social prescribing ontology, implemented according to the Web Ontology Language, was designed to cover several key concepts encompassing social determinants of health. Readv2 and Clinical Terms Version 3 codes were identified using the NHS Terms Browser. The Royal College of General Practitioners Research Surveillance Centre, a sentinel network of over 1000 primary care practices across England covering a population of more than 4,000,000 registered patients, was used for data analyses for a defined period (ie, January 2011 to December 2019).

Results: In all, 668 codes capturing social prescriptions addressing different social determinants of health were identified for the social prescribing ontology. For the study period, social prescribing ontology codes were used 5,504,037 times by primary care practices of the Royal College of General Practitioners Research Surveillance Centre as compared to 29,606 instances of use of social prescribing codes, including NHS England's recommended codes.

Conclusions: A social prescribing ontology provides a powerful alternative to the codes currently recommended by NHS England to capture detailed social prescribing activity in England. The more detailed information thus obtained will allow for explorations about whether outputs or outcomes of care delivery can be attributed to social prescriptions, which is essential for demonstrating the overall value that social prescribing can deliver to the NHS and health care systems.

(J Med Internet Res 2020;22(12):e23721) doi:10.2196/23721

KEYWORDS

social prescribing; clinical informatics; ontology; social determinants of health

Introduction

Approximately 80% of health outcomes are linked to social determinants of health, which include health-related behaviors as well as socioeconomic and environmental factors [1,2]. Social prescribing is a relatively recent initiative that has been developed to address the social determinants of health. National Health Service (NHS) England defines social prescribing as “a way of linking patients in primary care with sources of support within the community to help improve their health and wellbeing” [3]. Social prescriptions are varied and are mostly delivered by voluntary, community, and social enterprise (VCSE) organizations. The activities delivered by VCSEs range from health (eg, local walking groups), education (eg, dietary classes), skills development (eg, to facilitate employment), sports (eg, parkrun), and leisure or art (eg, singing groups) activities [4].

Despite its promise, a major barrier to the evaluation of social prescribing is the lack of data on what social prescribing activity is taking place and the outcomes delivered for people participating in these activities. This stems from the lack of information on the prescribed social prescriptions as well as variation in the quality of data recorded by clinicians [5].

In an attempt to address these gaps, NHS England worked with commissioners, practitioners, providers, evaluators, and other stakeholder groups to create a consensus Common Outcomes Framework (COF) on the outcomes and outputs that should be measured to demonstrate the impact of social prescribing. NHS England published the COF in 2019 [3] and recommended the use of 3 primary care codes to standardize the recording of social prescribing activity in primary care: “social prescribing offered,” “social prescribing declined,” and “referral to social prescribing service,” which are characterized as “finding’,” “situation,” and ”procedure,” respectively, in the Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) concept top-level hierarchy [6].

A standardization of how social prescriptions are recorded in primary care is essential to improve data quality so the general approach advocated by NHS England with the COF is sound. However, the codes recommended by NHS England have several limitations, which stem from their very general nature. For instance, the corresponding equivalent codes for pharmaceutical prescriptions would be “pharmaceutical prescription offered,” ”pharmaceutical prescription declined,” and ”pharmaceutical prescription given.” The general nature of these codes means that they do not capture the actual intervention delivered, which means that we cannot extrapolate which outcomes could realistically have been delivered by the social prescription; therefore, we cannot accurately attribute any outcomes to the actual social prescription. These limitations imply that if we only rely on these codes, it would be impossible to know whether social prescriptions deliver any benefit.

In this study, we explored whether an alternative approach to coding social prescribing activity can be used to capture more detail on the actual social prescriptions used. Specifically, we used well-established ontological approaches, which are used for modeling the semantics of medical concepts [7], to explore whether:

- a social prescribing ontology can be created with existing primary care codes to capture more detail on which social prescriptions are prescribed by primary care professionals
- the ontological codes are actually used by primary care professionals in practice
- a social prescribing ontology can serve as a viable alternative to capture more detailed information on social prescriptions in England

Methods

The study methods were essentially the same as previously reported [8] but are discussed briefly below.

Designing and Compiling the Ontology

An ontology is defined as a set of concepts and categories in a subject area or domain that describes their properties and the relations between them. The social prescribing ontology covers several key concepts derived from the “Five Ways to Wellbeing” model proposed by the New Economics Foundation [9] as well as Wilkinson and Marmot’s work [2] on social determinants of health (Figure 1 and Table 1).

Figure 1. Design of the social prescribing ontology.

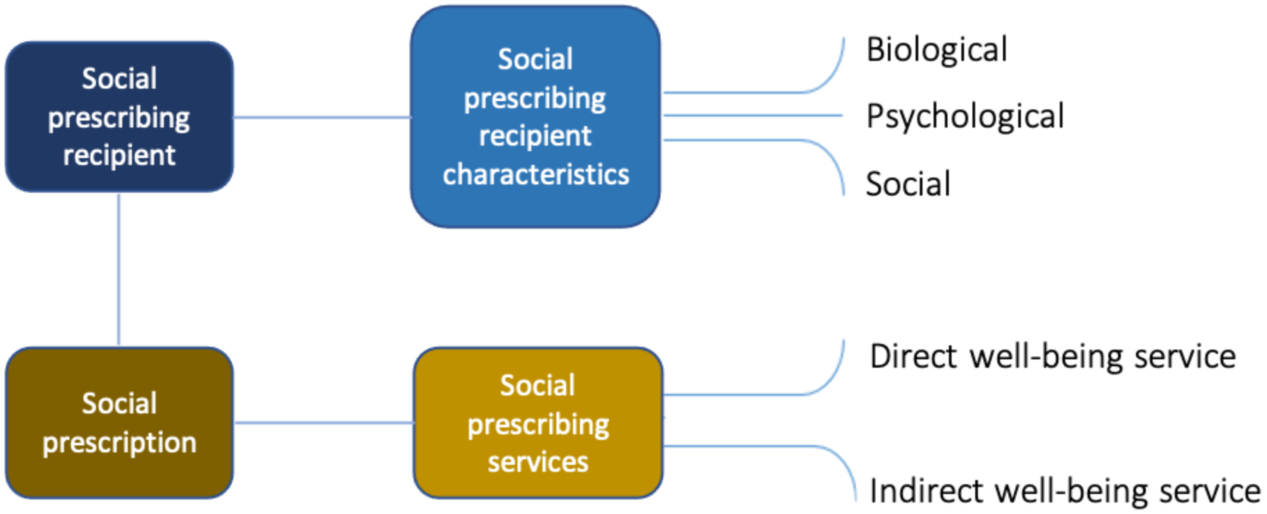


Table 1. Social prescribing ontological categories and unique primary codes (N=668) for each category.

Social prescribing ontological category	Unique primary care codes, n
Addictions support services	35
Benefits signposting services	10
Bereavement support services	20
Dementia support services	13
Diabetes management support services	11
Dietary support services	185
Domestic violence support services	2
Education support services	1
employment support services	20
finance support services	4
General lifestyle support services	15
General social support services	27
Home-based support services	19
Housing support services	25
Mental health services	17
Support services for other conditions	21
Parental support services	139
Physical activity management services	89
Respiratory support services	6
Stress reduction support services	9

The Readv2 and Clinical Terms Version 3 (CTV3) codes that comprise the social prescribing ontology were identified through 2 NHS Digital resources: (1) the NHS Term Browser, which is hosted by NHS Digital to provide a means to browse and search the SNOMED CT UK Edition, and (2) the Readv2 CTV3 to SNOMED CT Mapping Lookup, which maps SNOMED CT to the Readv2 and CTV3 terminologies. The social prescribing ontology has been implemented according to the Web Ontology Language (OWL) within the Protégé ontology development environment and hosted on the BioPortal ontology repository [10].

Data Analysis

We utilized the Royal College of General Practitioners Research Surveillance Centre (RCGP RSC) sentinel network as previously described [8]. The RCGP RSC was established in 1967 and comprises computerized medical record (CMRs) of pseudonymized data received from over 1000 primary care practices across England, covering a population of more than 4,000,000 currently registered patients [11,12].

CMR data in UK primary care centers are captured primarily within 2 electronic health record (EHR) systems that utilize Readv2 and CTV3 codes. Both these systems will be transitioning to SNOMED CT, but the analyses in this study relied on historical data from 2011 to 2019 so we did not use SNOMED CT codes in the data extracts. Readv2 and CTV3 codes are used to collate data for primary care, including diagnoses, processes of care, prescriptions, and results from laboratory-based data.

We extracted and analyzed coded, pseudonymized data from the RCGP RSC sentinel network primary care practices from January 1, 2011, to December 31, 2019. The data extracts included all instances of use of the codes highlighted in Supplementary Table S1 (Multimedia Appendix 1).

Ethical Approval

Consent was not required for the RCGP RSC data. Furthermore, data were not processed for individuals who had active opt-out codes present (which comprises 2.74% of registered patients as of March 7, 2019) [13]. The data were pseudonymized and encrypted before they were uploaded to the Clinical Informatics Research Group secure server. Personal data was not identifiable. This study was considered to be an "audit of current practice" when tested against the Health Research Authority/Medical Research Council "Is my study research" tool [14] and, therefore, did not require specific ethical approval. The RCGP RSC Study Approval Committee approved the use of data.

Data extractions were conducted in accordance with the Clinical Informatics and Health Outcomes Research Group's standard operating procedures for data extraction, pseudonymization, and transfer, as described previously [15].

Results

Social Prescribing Ontology

Twenty ontological categories were identified with a total of 668 codes heterogeneously distributed across all ontological categories, ranging from 185 codes for "Dietary support services" to only 1 code for "Education support services" (see Table 1 and Supplementary Table S1 in Multimedia Appendix 1).

Determining the Utilization of Social Prescribing Ontological Codes

The RCGP RSC dataset was searched from January 01, 2011 to December 31, 2019, to determine the extent to which codes within the social prescribing ontology were used by RCGP RSC primary care practices in England. Codes for "social prescribing," including the 3 codes recommended in the NHS England COF, were also investigated (for the full code list, see Supplementary Table S1 in Multimedia Appendix 1).

In all, 29,606 instances of use of "social prescribing" codes were found during the search period, compared to 5,504,037 instances of use of social prescribing ontology codes by RCGP RSC primary care practices (Table 2).

Table 2. Number of instances of use of social prescribing and social prescribing ontology codes within the Royal College of General Practitioners Research Surveillance Centre from January 01, 2011, to Dec 31, 2019 (N=5,533,643).

Category	Instances of code use recorded during the study period, n
Social prescribing codes	29,606
Social prescribing ontology code	5,504,037
Dietary support services	2,087,171
Physical activity management services	1,782,267
Addictions support services	769,860
General lifestyle support services	552,677
Parental support services	94,766
General social support services	75,321
Diabetes management support services	73,404
Homebased support services	22,198
Bereavement support services	16,212
Respiratory support services	9699
Support services for other conditions	7400
Mental health services	4868
Dementia support services	3710
Benefits signposting services	2169
Stress reduction support services	1012
Employment support services	743
Housing support services	554
Finances support services	6
Domestic violence support services	0
Education support services	0

Discussion

In this study, we found that a social prescribing ontology could be used to provide more details about the type of social prescription utilized by primary care practices in England. We identified 668 existing codes within Readv2 and CTV3 code sets that captured social prescriptions to a greater level of detail than those captured by the recommended NHS England codes of “social prescribing offered,” “social prescribing declined,” and “referral to social prescribing service.” We also found that the ontology codes were regularly used by primary care professionals across the nationally representative RCGP RSC sentinel network with over 5 million instances of use recorded between January 2011 and December 2019.

Our study demonstrates that primary care professionals have been regularly using the codes identified within our social prescribing ontology since 2011. This finding indicates these professionals were already aware of these codes and were using nonmedical interventions to address the social needs of patients through their existing primary care workforce, that is, before the establishment of link workers. With support from NHS England and key stakeholders, a social prescribing ontology could be recommended from a policy perspective, and it could be used nationally to improve data quality on social prescribing. Creating a national social prescribing ontology will be an iterative process that will require engagement with key stakeholders and consensus building—similar to the process used to create the COF. This process will also help clarify what can be truly characterized as a social prescription because some interventions such as education are not limited to only social prescribing, and this can ultimately inform the creation of new codes within SNOMED CT. Furthermore, given that the codes used for the ontology already exist in primary care code sets, templates could be created in primary care EHRs to facilitate access and utilization of these codes to more accurately capture social prescribing activity while also creating the digital infrastructure needed to create a social prescribing formulary [16].

Our study findings demonstrate that a social prescribing ontology, if appropriately designed, provides a powerful alternative to the codes currently recommended by NHS England to capture social prescribing activity. This is because such an ontology provides more granular information on the actual social prescription used, which will allow for explorations about whether outputs or outcomes of care delivery can be attributed to social prescriptions. These are essential steps for demonstrating the overall value that social prescribing can deliver to the NHS and health care systems.

Acknowledgments

AJ and HL were involved in creating the ontology, data analysis, and writing the manuscript. CO and UH were involved in creating the ontology and writing the manuscript. JS, FF, and IY were involved in data extraction. SdL was involved in preparing the manuscript. We thank Martin Cattermole and Else Krajenbrink for helpful discussions. The study was funding through a grant from NHS England.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Social prescribing ontology code list.
[DOCX File , 100 KB - jmir_v22i12e23721_app1.docx]

References

1. Hood C, Gennuso K, Swain G, Catlin B. County health rankings: relationships between determinant factors and health outcomes. Am J Prev Med 2016 Feb;50(2):129-135. [doi: 10.1016/j.amepre.2015.08.024] [Medline: 26526164]
2. Wilkinson R, Marmot M, editors. Social determinants of health: The solid facts. Second edition Internet. Denmark: World Health Organization; 2003.
3. Social prescribing and community-based support: Summary guide. NHS England. 2019 Jan 31. URL: https://www.england.nhs.uk/publication/social-prescribing-and-community-based-support-summary-guide/ [accessed 2020-08-17]
4. Jani A, Gray M. Making social prescriptions mainstream. J R Soc Med 2019 Nov;112(11):459-461. [doi: 10.1177/0141076819848304] [Medline: 31710824]
5. Bickerdike L, Booth A, Wilson PM, Farley K, Wright K. Social prescribing: less rhetoric and more reality. A systematic review of the evidence. BMJ Open 2017 Apr 07;7(4):e013384 [FREE Full text] [doi: 10.1136/bmjopen-2016-013384] [Medline: 28389486]
6. SNOMED CT Concept Model. SNOMED International Home. URL: https://confluence.ihtsdotools.org/display/DOCSTART/6.+SNOMED+CT+Concept+Model [accessed 2020-09-20]
7. Liyanage H, Krause P, De Lusignan S. Using ontologies to improve semantic interoperability in health data. J Innov Health Inform 2015 Jul 10;22(2):309-315 [FREE Full text] [doi: 10.14236/jhi.v22i2.159] [Medline: 26245245]
8. Jani A, Liyanage H, Hoang U, Moore L, Ferreira F, Yonova I, et al. Use and impact of social prescribing: a mixed-methods feasibility study protocol. BMJ Open 2020 Sep 18;10(9):e037681 [FREE Full text] [doi: 10.1136/bmjopen-2020-037681] [Medline: 32948564]
9. Aked J, Marks N, Cordon C, Thompson S. Five Ways to Wellbeing: A report presented to the Foresight Project on communicating the evidence base for improving people’s well-being. Repository for Arts & Health Resources.: New Economics Foundation; 2008. URL: https://neweconomics.org/uploads/files/8984c5089d5c2285ee_t4m6bhqq5.pdf [accessed 2020-12-03]
10. Social Prescribing Ontology. Bioportal. 2018 Dec 10. URL: https://bioportal.bioontology.org/ontologies/SOCPRES [accessed 2020-09-25]
11. de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, et al. RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. Br J Gen Pract 2017 Oct;67(663):440-441 [FREE Full text] [doi: 10.3399/bjgp17X692645] [Medline: 28963401]
12. Correa A, Hinton W, McGovern A, van Vlymen J, Yonova I, Jones S, et al. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. BMJ Open 2016 Apr 20;6(4):e011092 [FREE Full text] [doi: 10.1136/bmjopen-2016-011092] [Medline: 27098827]
13. [MI] National Data Opt-out, March 2019. NHS Digital. 2019 Mar 19. URL: https://digital.nhs.uk/data-and-information/publications/statistical/national-data-opt-out/march-2019/ndop-mar19 [accessed 2020-08-17]
14. Is my study research? NHS Health Research Authority. URL: http://www.hra-decisiontools.org.uk/research/ [accessed 2020-08-17]
15. de Lusignan S, Borrow R, Tripathy M, Linley E, Zambon M, Hoschler K, et al. Serological surveillance of influenza in an English sentinel network: pilot study protocol. BMJ Open 2019 Mar 08;9(3):e024285 [FREE Full text] [doi: 10.1136/bmjopen-2018-024285] [Medline: 30852535]
16. Jani A, Pitini E, Jungmann S, Adamo G, Conibear J, Mistry P. A social prescriptions formulary: bringing social prescribing on par with pharmaceutical prescribing. J R Soc Med 2019 Dec;112(12):498-502. [doi: 10.1177/0141076819877555] [Medline: 31825285]

Abbreviations

CMR: computerized medical record
COF: Common Outcomes Framework
CTV3: Clinical Terms Version 3
EHR: electronic health record
NHS: National Health Service
OWL: Web Ontology Language
RCGP RSC: Royal College of General Practitioners Research Surveillance Centre
SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms
VCSE: voluntary, community and social enterprise

Edited by R Kukafka; submitted 20.08.20; peer-reviewed by S Madani, C Bonner; comments to author 17.09.20; revised version received 27.09.20; accepted 19.10.20; published 11.12.20.

Please cite as:
Jani A, Liyanage H, Okusi C, Sherlock J, Hoang U, Ferreira F, Yonova I, de Lusignan S
Using an Ontology to Facilitate More Accurate Coding of Social Prescriptions Addressing Social Determinants of Health: Feasibility Study
J Med Internet Res 2020;22(12):e23721
URL: http://www.jmir.org/2020/12/e23721/
doi:10.2196/23721
PMID:33306032

Original Paper

Status and Recommendations of Technological and Data-Driven Innovations in Cancer Care: Focus Group Study

Haridimos Kondylakis[1*], BSc, MSc, PhD; Cristian Axenie[2*], PhD; Dhundy (Kiran) Bastola[3*], PhD, MSc; Dimitrios G Katehakis[1*], MSc; Angelina Kouroubali[1*], PhD; Daria Kurz[4*], PhD; Nekane Larburu[5*], PhD; Iván Macía[5*], PhD; Roma Maguire[6*], PhD; Christos Maramis[7*], PhD; Kostas Marias[1*], PhD; Philip Morrow[8*], PhD; Naiara Muro[5*], PhD; Francisco José Núñez-Benjumea[9*], PhD; Andrik Rampun[10*], PhD; Octavio Rivera-Romero[11*], PhD; Bryan Scotney[8*], BSc (Hons), PhD; Gabriel Signorelli[9*], PhD; Hui Wang[12*], PhD; Manolis Tsiknakis[1*], PhD; Reyer Zwiggelaar[13*], PhD

[1]FORTH-ICS, Heraklion, Greece
[2]Audi Konfuzius-Institut Ingolstadt Lab, Technische Hochschule Ingolstadt, Ingolstadt, Germany
[3]School of Interdisciplinary Informatics, University of Nebraska, Omaha, NE, United States
[4]Interdisziplinäres Brustzentrum, Helios Klinikum München West, Munich, Germany
[5]Vicomtech, Health Research Institute, San Sebastian, Spain
[6]University of Strathclyde, Glasgow, United Kingdom
[7]eHealth Lab, Institute of Applied Biosciences - Centre for Research & Technology Hellas, Thessaloniki, Greece
[8]School of Computing, Ulster University, Newtownabbey, United Kingdom
[9]Salumedia Labs & Adhera Health, Inc, Sevilla, Spain
[10]Academic Unit of Radiology, Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, United Kingdom
[11]Universidad de Sevilla, Seville, Spain
[12]School of Computing and Engineering, University of West London, London, United Kingdom
[13]Department of Computer Science, Aberystwyth University, Aberystwyth, United Kingdom
[*]all authors contributed equally

Corresponding Author:
Haridimos Kondylakis, BSc, MSc, PhD
FORTH-ICS
N Plastira 100
Vassilika Vouton
Heraklion, 70013
Greece
Phone: 30 2810391449
Email: kondylak@gmail.com

Abstract

Background: The status of the data-driven management of cancer care as well as the challenges, opportunities, and recommendations aimed at accelerating the rate of progress in this field are topics of great interest. Two international workshops, one conducted in June 2019 in Cordoba, Spain, and one in October 2019 in Athens, Greece, were organized by four Horizon 2020 (H2020) European Union (EU)–funded projects: BOUNCE, CATCH ITN, DESIREE, and MyPal. The issues covered included patient engagement, knowledge and data-driven decision support systems, patient journey, rehabilitation, personalized diagnosis, trust, assessment of guidelines, and interoperability of information and communication technology (ICT) platforms. A series of recommendations was provided as the complex landscape of data-driven technical innovation in cancer care was portrayed.

Objective: This study aims to provide information on the current state of the art of technology and data-driven innovations for the management of cancer care through the work of four EU H2020–funded projects.

Methods: Two international workshops on ICT in the management of cancer care were held, and several topics were identified through discussion among the participants. A focus group was formulated after the second workshop, in which the status of technological and data-driven cancer management as well as the challenges, opportunities, and recommendations in this area were collected and analyzed.

Results: Technical and data-driven innovations provide promising tools for the management of cancer care. However, several challenges must be successfully addressed, such as patient engagement, interoperability of ICT-based systems, knowledge management, and trust. This paper analyzes these challenges, which can be opportunities for further research and practical implementation and can provide practical recommendations for future work.

Conclusions: Technology and data-driven innovations are becoming an integral part of cancer care management. In this process, specific challenges need to be addressed, such as increasing trust and engaging the whole stakeholder ecosystem, to fully benefit from these innovations.

(J Med Internet Res 2020;22(12):e22034) doi:10.2196/22034

KEYWORDS

neoplasms; inventions; data-driven science

Introduction

Background

The morbidity and mortality associated with cancer are rapidly increasing globally because of population growth and aging, reflecting the changes in the prevalence and distribution of major risk factors of cancer [1]. These global trends have resulted in more people living with or beyond cancer. As such, there is a greater need to improve and optimize cancer care services throughout diagnosis, treatment, rehabilitation, and end-of-life care. Modern technologies, often enabled by the availability of big data and advanced analytics, have demonstrated the potential to enhance the current level of quality of cancer care, for example by improving information access, informing and sharing clinical decision making with patients, and facilitating communication and support for coliving with the illness.

Pioneering research is now being conducted in the field of cancer care technology, resulting in the development of novel solutions to a diverse spectrum of problems in this area. However, the process of evaluating these innovations and their operation within a *real-world* context is at a less advanced stage. The clinical assessment of technology poses many challenges and is influenced by numerous variables. As a result, there is an emphasis on the early evaluation of technology by including key stakeholders throughout the design and development phases (sometimes referred to as a cocreation process).

Objectives

Evidently, as researchers involved in the creation of novel technologies for cancer care, we must consider and share both the innovative concepts being developed as well as how they are being assessed and accepted in real-world or clinical settings.

To this end, an international workshop was convened to consider the current status of technological and data-driven innovations in cancer care, to identify key challenges and opportunities, and to formulate recommendations aimed at accelerating the rate of progress in the data-driven management of cancer. The two workshop instances led to a series of publications. This paper discusses key topics arising from the workshops and subsequent discussions among their participants.

Methods

An international expert consensus-building workshop named Tech4Cancer [2] was held in two instances; the first was held in June 2019 in Cordoba, Spain, along with the computer-based medical systems [3] international conference, and the second was held in October 2019 in Athens, Greece, along with the Institute of Electrical and Electronics Engineers (IEEE) Bioinformatics and Bioengineering [4] international conference. The workshops attracted significant international participation; more than 40 participants attended these events, where they focused on discussing global trends for technological and data-driven innovations in cancer care.

The workshop was supported by four European Union (EU) projects, namely BOUNCE, CATCH ITN, DESIREE, and MyPal, funded by Horizon 2020 (H2020). The BOUNCE project [5,6] considers clinical, cancer-related biological, lifestyle, and psychosocial parameters to predict individual resilience trajectories throughout the cancer continuum. Eventually, the target is to increase resilience in breast cancer survivors and help them remain in the workforce and have a better quality of life. DESIREE [7], on the other hand, aims to provide a web-based software ecosystem for personalized, collaborative, and multidisciplinary management of primary breast cancer by specialized breast units, from diagnosis to therapy and follow-up.

MyPal [8] aims to foster palliative care for people with cancer by leveraging patient-reported outcome (PROM) systems through their adaptation to the personal needs of the person with cancer and his or her caregivers [9]. In this regard, MyPal designed two novel eHealth interventions based on electronic PROMs to offer advanced palliative care services to adult patients with cancer and children with cancer. MyPal interventions are delivered via a sophisticated eHealth platform developed by the project. The interventions and the platform itself will be assessed through two multicenter clinical studies (one for the case of adults and one for the case of children), which will take place throughout Europe. CATCH ITN [10] is a PhD training network, with academic and industry partners across Ireland, Spain, and Denmark. Research within CATCH ITN focuses on the development and use of technology to improve the quality of life of individuals living with and beyond cancer.

The primary deliverable from this workshop was a set of articles that summarize key issues already published by IEEE. During the paper presentation, several topics were identified through discussion among the participants. After the second instance of the workshop, participants were invited to participate in a focus group discussing status, challenges, opportunities, and recommendations in the area of technological and data-driven cancer management; the outcome is reported in this study.

The writing group met by teleconference, and participants were asked to propose topics according to their interest and expertise and to select topics that they could actively contribute to. Several topics depicting challenges, opportunities, and recommendations were selected. Subsequently, leaders for each topic were identified and a structure was proposed. A formal consensus process was not used; however, the structured and open discussions did not reveal any fundamental disagreements about the nature of the topics, although the discussions supported the refinement and specificity of topics. Once all contributions were collected, a homogenization and integration process led to a final draft, where all participants commented and discussed, leading to the final submitted version.

Results

Patient Engagement and Participatory Design

A fundamental requirement for the effectiveness of any eHealth intervention, including interventions addressed to patients with cancer and survivors, is a certain level of *engagement* by the patient [11], who is the main beneficiary of the intervention [12]. The lack of engagement of patients with eHealth interventions is linked to low motivation and adherence to the intervention, leading to high dropout rates and eventually affecting treatment outcomes and effectiveness [13]. The issue of patient engagement has cast a shadow over the enthusiasm for the potential of eHealth [14].

Formulating a concrete definition of patient engagement in the context of eHealth is a challenge by itself [14], as there is a lack of consensus on what the term engagement entails and because different fields (eg, computer science, psychology, and behavioral health) conceptualize it differently. A working definition comes from the Canadian Institutes for Health Research in the framework of the Strategy for Patient-Oriented Research: patient engagement is the meaningful and active collaboration of the patient in governance, priority setting, research, and knowledge translation [15]. Some researchers have defined engagement in terms of the level of *activation* of the patient in the management of their own health [16]. Certain definitions rely on psychological processes related to user perceptions and experience, whereas others consider engagement as a purely behavioral construct, synonymous with intervention usage [17]. Consequently, engagement is often used interchangeably with *adherence*, which refers to whether the intervention is used as intended by its developers [18,19]. This is the definition of patient engagement that is primarily considered in the context of this study.

The situation is not significantly better when it comes to the assessment of patient engagement. The lack of consensus in the conceptualization of user engagement makes the design of appropriate *universal* or even *widely applicable* assessment instruments difficult. Instead, most research efforts propose solutions for assessing user engagement for very specific purposes, modalities, and contexts pertaining to eHealth [11,20,21]. Focusing on mobile health (mHealth), a prominent subfield of eHealth, a recent scoping review analyzed 41 studies and developed a library of 7 analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions, namely the (1) number of measures recorded, (2) frequency of interactions logged, (3) number of features accessed, (4) number of log-ins or sessions logged, (5) number of modules or lessons started or completed, (6) time spent engaging with the app, and (7) number or content of pages accessed [22]; this work is of particular relevance to care for cancer, which currently tends to be considered a chronic disease [23].

The scientific literature sheds some light on the techniques that have been employed to achieve, maintain, and improve patient engagement. According to a psychology study [24], patient engagement techniques are categorized as follows:

1. Behavioral techniques, such as motivational interviewing, goal setting, and planning, which are related to patient actions when managing their health condition.
2. Cognitive techniques, such as question-asking tasks and psychoeducation sessions, which are related to patient thoughts and received information concerning their health condition.
3. Emotional techniques, such as positive psychology exercises and expressive writing tasks, which are related to experienced patient feelings and emotions when adjusting to their new health condition.

Most interventions developed for older adults employ patient engagement techniques from behavioral and/or cognitive categories, but not all categories [24]. The latter is also the conclusion of a systematic review of eHealth for patient engagement [25]. Moving beyond the psychology-oriented categorization, other patient engagement techniques, such as shared decision making and brief negotiated interviewing, have also been employed; in fact, interventions have been designed to integrate both techniques for engagement optimization [26]. Personalization has also been adopted in some studies as a patient engagement technique [16,27].

However, the main barrier in building a critical mass of literature on patient engagement with eHealth systems is the fact that very few studies address or report the topic. For instance, a systematic review on published trials discovered that only 23 of 2777 reviewed trials reported any patient engagement activities [28]. The literature review that was conducted in the framework of a different study focusing on older patients concluded that interventions to engage patients are generally scant and often poorly described [24]. In addition, in a recent systematic and mapping review on eHealth interventions employing electronic patient-reported outcomes (ePROs) for palliative cancer care conducted in the framework of the MyPal project [29], 17 of

24 eligible ePRO-based palliative care interventions for patients with cancer did not take patient engagement into account in the development or evaluation of the proposed digital health intervention. This may have been one of the reasons for the high attrition rates reported by the studies. Furthermore, four other studies recognized—mostly retrospectively—the need for patient engagement, and some attempted to measure it. Only 3 remaining studies actively attempted to optimize patient engagement during the design of the eHealth intervention or system.

Opportunities

In contrast to the other challenges discussed in this work, the solution to patient engagement may not be rooted in technology itself but rather related to the way technology is designed. To this end, there is a growing body of research demonstrating the value of cocreation and participatory design in the development of novel digital health services, including services for cancer care. Participatory design is one of the pillars of the revolutionary predictive, personalized, preventive, and participatory (P4) cancer medicine [30,31], and it requires the active involvement of users in the design or development phase of an eHealth system. Participatory design can provide a unique perspective on user acceptability, system usability, and the feasibility of the overall effort [32]. By involving a representative sample of a population of patients with cancer in the design or development of an eHealth system or service, one can eventually build an innovative solution that is expected to have good engagement with the target population.

The main barrier is a lack of research culture for considering the involvement of the end user (ie, in our case, the patient with cancer) in the design process. Despite this barrier, participatory design is expected to become the norm in eHealth technology development in the upcoming years.

Recommendations

The MyPal project (see the *Methods* section) is an example of a research effort that has committed to a *participatory design approach*, implemented primarily via a series of focus group discussions on eHealth interventions to be developed with the participation of all the involved stakeholders (patients as well as their informal carers and treating health care professionals) [33].

The use of the participatory approach as early as possible in the design of innovative technological solutions for patients with cancer presents a good opportunity to improve patient engagement [34]. The methodological tools, coming mostly from the field of qualitative research (eg, semistructured focus group discussions), are mature enough to support this design paradigm in eHealth, as these have been extensively validated in more generic software engineering environments. The employed participatory design process has driven the development of the MyPal platform for palliative cancer care and has provided a series of generalizable guidelines or recommendations for the successful application of participatory design processes for patients with cancer engagement, which are as follows:

1. Participatory design should start as early as possible in the development lifecycle of an eHealth system or service, and it should rely on established methodological tools.
2. Representative samples of the intended patient populations need to be selected for participation in the co-design and cocreation activities. This is especially important for heterogeneous patient populations.
3. Participatory design findings should be fused with other sources of knowledge (eg, a screening of unmet patient needs from cancer care in the scientific literature).

These recommendations can complement pre-existing published efforts to deliver more generic guidelines for developing engaging eHealth technologies. For example, the work presented in Karekla et al [14] specified 10 recommendations for researchers and clinicians interested in developing an engaging eHealth system. These recommendations are organized around 4 dimensions, namely a priori theoretical planning, human-computer interaction, tailoring and targeting to user groups, and active assessment of use. The eHealth technology development framework presented in Gemert-Pijnen et al [35] can also serve as a source of akin recommendations. Although the main goal of the framework is to improve eHealth technology uptake, some of the 6 working principles it introduces are applicable to the pursuit of patient engagement, especially the principle advocating for persuasive design techniques.

Small Data Analytics

Data are fuel for any machine learning (ML) project [36]. It is well known that deep supervised learning algorithms are particularly data hungry—not only do they need a lot of data samples but the data also have to be manually annotated beforehand. However, big data sets with annotations (labeling, structuring, etc) are very rare, as proper annotations must be done by experts, and this is very expensive. Therefore, annotated data sets are mostly small data. Annotating data in a less costly manner remains to be a key challenge. A related challenge is to reduce the dependence of ML on annotation.

ML, especially deep learning, can effectively learn with big data. However, it cannot effectively learn with small data because of various issues, for example, overfitting, noise, outliers, and sampling bias, which can render the learned model effectively useless. Effective learning with small data is a challenge.

Opportunities

The annotation problem can be addressed via the use of annotation tools or services in many ways, such as (1) providing annotation tools so that annotation can be performed more effectively and easily, existing tools include Lionbridge artificial intelligence (AI) [37] and Computer Vision Annotation Tool [38]; (2) outsourcing the annotation task to an annotation service provider such as Amazon Mechanical Turk [39]; and (3) enabling an expert to teach their ML model while building and annotating their data set. However, none of these approaches solve the annotation problem scientifically. There are ways to deal with learning with small data problems, such as data augmentation, transfer learning, regularization, and visualization.

However, these methods require skilled people, and their effectiveness is limited.

Another research question is regarding the procedure to add the value of ML results under the constraint of the available knowledge. Moreover, as knowledge and the latest clinical evidence, such as clinical practice guidelines (CPGs), are usually in paper-based formats and written in natural language, another research question arises regarding the procedure to automatically represent knowledge in a structured and computerized manner.

Recommendations

To solve the annotation problem scientifically, a desirable approach is to design a new ML algorithm that requires minimal feedback from human experts. This will only be possible if domain-specific constraints can be imposed on the learning process. This will reduce the model space as well as the variance in learning. Thus, one research question is on the procedure to reduce model space by domain-specific constraints.

An interesting approach to solve the *learning with small data* problem is to use domain knowledge in the learning process or *knowledge-based learning*. ML requires data as well as knowledge (common sense and domain-specific knowledge) implicitly or explicitly. When there are a lot of data, ML requires a small amount of knowledge; when there is not much data, ML requires a large amount of knowledge to reduce the model search space. Model-based ML can be seen as an example of knowledge-based learning, where knowledge can be specified by experts in the form of variables and their dependencies. This approach has been successfully demonstrated in various case studies. One weakness, however, is that formalizing knowledge is not a straightforward task, and it requires capturing experience from clinicians through inverse engineering and making clinical statements as explicit as possible for a computerized system. A desirable approach is a knowledge-based learning algorithm that (1) can be easily extended by data-driven findings and (2) uses standardized terminologies to provide interoperability and eases the updating and maintenance of the latest evidence. For example, within the DESIREE project, a digital breast cancer patient was formalized as a knowledge model ontology. The ontology employed standardized terminologies to identify univocally all identified clinical terms and procedures, which included the knowledge reported in standardized guidelines such as the National Comprehensive Cancer Network (NCCN) [40] or the European Society for Medical Oncology [41] guidelines in a computerized format, to be provided as recommendations through a decision support system (DSS) for multidisciplinary tumor boards during the decision-making process. Although formalization of knowledge through a model is almost a mandatory task in the very first steps of technical developments, it is a very time-consuming and costly task.

Integration and Data Management

The multitude of successful instantiations of digital interventions has opened exciting new directions for acquiring, delivering, and sharing data, and has already proven the potential to leverage cost-effective, patient-centered cancer care applications [42-44]. Embedding patients into the iterative design process of digital intervention has been shown to enable developers to increase the relevance and effectiveness of the intervention [42]. Such treatments should be created with the acknowledgment of patients’ knowledge, attitudes, beliefs, preferences, and expectations of therapeutic outcome [45,46]. Thus, systems that support clinicians would benefit from a redesign that aligns cancer care more completely with patients’ needs and interests [47,48]. ML, through its pervasive impact [49], has the potential to provide a supportive tool for such a redesign task, given the powerful data understanding, generalization capabilities, and robustness [50].

Opportunities

Guided by the current state of the art, we identified a set of opportunities for digital interventions related to patient-centered cancer care. We have considered the most cited systematic reviews in the last 3 years (ie, to extract the innovation opportunities and technological limitations) and older systematic reviews (ie, studies from 2011 to tackle the initial adoption and strategies). These opportunities tackle the inherent challenges that we identified in practice. The first challenge is data collection and integration. Here, the core opportunity relates to exploring and exploiting the aggregation of heterogeneous and distributed data, including personal, professional, and health-related information [51]. Such systems are typically deployed as a platform to integrate retrospective, prospective, and day-to-day care data. However, data interpretation and translation remain to be an open challenge, whereas there is also a constant need for standard data formats such as Digital Imaging and Communications in Medicine and other standards applicable to electronic patient record or electronic health record (EHR) data. To address this, there is a clear opportunity to interpret patient information for the patient themselves by translating the clinical findings into an intuitive representation. This would ensure that the delivery of health information is done via a user-friendly interface.

When considering clinical data heterogeneity and the specificity of patient data, data projection is another challenge that is present in any technological intervention. We delineate a good opportunity to build individualized clinical recommendations based on the data interpretation projected in actions upon the patient. This initiative will be based on the patient’s risk profile and evidence-based guidelines (see the *Extracting Patient Portraits* section).

Finally, with regard to data sparsity, such clinical setups face the challenge of data completeness and augmentation. In contrast to clinical trial experience, data completeness improves with longitudinal care. This approach may be a solution to minimizing missing data of PROMs in research or clinical care settings in support of learning health care systems capable of augmenting the data [52]. To maintain the validity of the intervention, we believe that there is a good opportunity to consider and consolidate clinical data with the comprehension of lived experiences of patients, centered on patients [53].

Recommendations

On the basis of the identified challenges and opportunities, we established a set of recommendations for the systematic development of appropriate patient-centered digital interventions that ensure usefulness, adoption, and sustainability in cancer care [54,55]. This process can be extended with learning and generalization capabilities. ML algorithms excel at such tasks and constitute invaluable tools for any digital intervention.

Next, we map the challenges that we identified before practical recommendations. These recommendations can serve as a reference design for technological interventions when looking at integration and data management.

When addressing data collection and integration, we recommend embedding continuous user feedback and iterative prototyping in the intervention. This can be achieved by exploiting the multimodal nature of patient data (ie, personal, behavioral, professional, and health-related). ML techniques (ie, deep learning and hybrid neural networks) can fuse heterogeneous data in a common representation (ie, efficiently using very large data sets containing health care use data, clinical data, and data from personal devices and many other sources), as demonstrated by the recent deep learning systems used on multi-omics data sets to drive precision oncology care [56].

In terms of data interpretation and translation, we recommend the use of tools to extract and represent the medical substrate by synthesizing only relevant aspects in a declarative way. ML techniques (ie, deep learning—recurrent networks with word embeddings and distributed representations) can handle very large and sparse data (eg, device data may only be available for a small subset of individuals) to capture the sequential character of the data and are suitable for modeling context dependencies in inputs [57]. Such systems, which incorporate word embeddings encoding syntactic and polarity information in the language followed by deep neural network architectures, are already used to extract and normalize parameters within oncology care data.

To address the challenges in the area, we recommend the development of clinical projections (ie, mappings) from individualized patient recommendations to therapy plans that embed temporal, procedural, and reasoning processes. ML techniques, such as temporal hierarchical task networks, can dynamically generate personalized therapy plans for oncology patients [58], following a deliberative hierarchical planning process driven by the procedural knowledge described in oncology protocols [59]. Such instantiations use mappings to attach reasoning and procedural knowledge representation as well as their interpretation in a temporal planning process. The planning process allows us to obtain temporally annotated therapy plans that support decisions of oncologists. Moreover, such an ML technique offers the ability to deal with complex temporal and resource constraints, typical in cancer care.

Finally, tackling data augmentation, we recommend incorporating the lived experiences of the patients [60]. ML techniques, such as contextualized word embeddings, are suitable for improved text augmentation independent of any task-specific knowledge or rules and can process structured questionnaires for patients who, for instance, developed chemotherapy-induced peripheral neuropathy [61].

Extracting Patient Portraits

Individual biomedical and nonbiomedical patient characteristics should guide any provided chronic care—digital or not. These insights are used to develop and validate *patient portraits* that can be employed in practice to determine optimal treatment strategies for subgroups of patients with similar cancer care needs and preferences. Building a *patient portrait* is hence an endeavor that follows a bottom-up approach that includes *patient profiling* based on patient phenotype algorithms [62-64]; intelligent patient profiling for the decision support of cancer treatment by exploiting clinical and genomic data [65]; personalized predictive modeling and risk factor identification using patient similarity [66]; and individualization beyond biomedical factors to also include demographic, socioeconomic, and psychological aspects [67]. To build a complete description of the multiple dimensions describing each patient, we identified a series of opportunities that describe all previously reviewed work.

Opportunities

The series of opportunities we identified target the selection of relevant heterogeneous and multimodal data correlated with the diagnosis. To extract the opportunities and frame our recommendations, we systematically analyzed a series of studies ranging from genomic data and phenotypes to demographics and psychological data. This methodology allowed us to capture the most relevant dimensions for extracting a patient profile. More precisely, the focus was on exploiting the correlations among multiple data sources to build a digital patient portrait consistent along all dimensions.

Such an initiative requires powerful data mining and ML algorithms, which can provide an efficient and compressed representation of a patient's digital profile, subsequently guiding therapeutic schemes.

One challenge identified relates to data relevance. Here, there is an opportunity to identify relevant genetic, phenotypical, physiological, lifestyle, and medical data correlated with the diagnosis. Exploiting such an opportunity can improve a patient's profile and the overall effect of the intervention, especially in progression-free survival [68]. Such studies demonstrate the feasibility of intelligent patient profiling that can select, within a clinically relevant time frame, a beneficial treatment for patients with no other treatment options.

Given the data deluge describing each patient, exploring and exploiting data correlations is another challenge we identified in the context of extracting a patient’s profile. This challenge provides a clear opportunity to exploit the correlations among the multiple identified modalities toward building an individual or personalized patient digital portrait that consistently captures all dimensions of the patient’s disease evolution.

Such an exploration unveils another challenge, namely multimodal data fusion [69], and the opportunity and potential that fusing extracted knowledge have toward a personalized therapeutic scheme [70]. Such systems, by combining data describing complementary perspectives on the same biological phenomena, can (1) separate correlated from discordant data,

(2) extract the most informative features, and (3) estimate disease progression.

The last challenge we identified as being crucial in any effort to extract a patient's profile is the possibility of embedding individualized data (eg, age, gender, ethnicity, health conditions, and social position) in patient cohorts [71]. Initial efforts were made to include family history for risk assessment and early detection of cancer; however, adherence to the study was low because of the limited technological support. This challenge brings along the opportunity to determine the population of interest and use this information in the process of *portrayal*.

Recommendations

On the basis of the identified opportunities and related challenges, we established a set of recommendations supporting a digital intervention design that (1) exploits available data, (2) extracts underlying correlations, and (3) integrates the multitude of representations in a structured object guiding therapeutic interventions in cancer care.

When addressing the challenge of identifying data relevance, we recommend selecting the data or feature subset that best characterizes the statistical property of a certain variable (ie, a certain patient) subject to the constraint that these data or features are as mutually dissimilar to each other as possible but as marginally similar to a certain class of patients as possible. For this task, ML tools such as minimum redundancy feature selection (ie, Minimum Redundancy Maximum Relevance) [72]) can be used to accurately identify the characteristics of patient features or data and narrow down their relevance. Such techniques provide an integrative approach to patient-centered data and demonstrate the potential of feature selection in data analysis and predictive patient-specific outcomes [73].

To address data correlations, we recommend the use of ML models, such as long short-term memory (LSTM) [74] networks, which can be used for their ability to effectively model varying length time series data and capture long-term dependencies and correlations. From modeling the patient life expectancy from medical records [75] to complete patient trajectories [76], modeling the disease trajectory and care processes, assumes mining electronic medical records that are episodic and irregular in time. Such models capture long-term temporal dependencies and are well suited to modeling clinical data because the evidence of certain conditions may be spread apart over several hours or days, and important symptoms may present early on in a patient's trajectory.

The next challenge identified is the fusion of the available multimodal data. Our recommendation is to use a data-driven feature learning class of approaches. Typically, they are based on deep networks that can directly learn the hidden characteristics of the data from different sources. As such, we recommend, for instance, the use of deep neural networks to extract features from genomic and clinical data [77], convolutional neural networks to extract features from pathology images [78], and recurrent neural networks for text and medical records data [79].

Finally, the last identified challenge is the possibility of embedding individualization data in the patient profile. We recommend performing individual cognitive interviews and focus groups with patients to learn about their relevant needs, experiences, fears, aspirations, and expectations.

From the ML point of view, a solution for developing personalized patient embeddings that is capable of processing such data is a combination of well-proven autoencoder methods with extensions to some of the metrics to account for data sparsity and multimodality [80]. Such studies also provide a methodology describing how these networks can be designed, built, and applied to tasks of integrative analyses of heterogeneous cancer data.

A patient portrait that can capture complex relationships in physiological signals, nonbiomedical data, and personal data embedding is key to accurately predict the stages of interventions for different patients and is necessary for successful personalized therapy.

Learning Patient Disease Trajectory for Personalized Diagnosis

Cancer is remarkably heterogeneous across individuals. This heterogeneity makes treatment difficult for caregivers because they cannot accurately predict how the disease will progress to guide treatment decisions. Therefore, tools that help to predict the individualized trajectory of cancer can help improve the quality of health care [81,82]. Given the assessment of the current state of the patient (ie, patient profile, physiology, neuroimaging, blood biomarkers, and physiologic testing) along with the therapeutic scheme, a digital intervention would use ML or predictive systems to infer disease evolution or remission to be able to guide subsequent therapy scheme planning [83]. Such a trajectory can also support the detection of behavior change in patients [84].

Opportunities

A significant need relates to making disease predictions by leveraging baseline information and additional time-dependent clinical markers as they are collected. Such an approach is the focal challenge of personalized medicine: integrative analysis of heterogeneous data from an individual's medical history to improve cancer care. We identified several key challenges and associated opportunities linked to this.

The first challenge relates to the fact that markers in clinical data are irregularly and sparsely sampled. Here, we identified a valuable opportunity for handling data and choosing specific latent variable models to summarize and extract information from the irregularly sampled and sparse data. This should simultaneously ensure sidestepping the issue of jointly modeling the data-generating processes [85]. Such systems should build a temporal representation of care trajectories in the form of a time-ordered state sequence. Moreover, in addition to the routine identification of key dates and events in patients' care trajectories, such systems should identify initially fragmented data across numerous sources.

Another challenge identified relates to the learning of a disease trajectory and is linked to the inherent computational complexity. Imposed by the clinical setup, we identified an opportunity to predict the entire disease progression trajectory from the observed patient records without many training labels on the ground-truth stages that a patient acquired, in mechanistic models of disease progression. A joint approach is prone to alleviate the inherent variability in prediction.

This challenge opens the stage for the next challenging point, namely continuous adaptation and updates in the face of disease progression heterogeneity. In handling such a challenge, there is a clear opportunity for continuous-time adaptation and updates to new observations and new data (markers). This provides novel computational methods for predicting, for instance, disease phenotype from molecular and genetic profiling [86].

Finally, another challenge we identified refers to the observed versus latent data artifacts. Addressing such a challenge demands tools for capturing latent factors in disease expression and not only observed features as a crucial aspect for patient-tailored cancer therapies. We further elaborate this in the following *Recommendations* section.

Recommendations

In this section, we match the challenges to practical recommendations in designing digital interventions when predicting disease trajectories for patients with cancer.

To handle irregularly and sparsely sampled markers in clinical data, we recommend, from the ML perspective, the use of discriminative models that condition on marker histories instead of jointly modeling them. Such an approach will not be sensitive to miss-specified dependencies across marker types and inherent irregularities and sparsity. For example, functional data analysis [87] can be employed for sequences of measurements that are assumed to be samples from an underlying continuous function. However, coefficient estimates can have a high variance in time series.

However, the task of predicting the disease trajectory comes with its inherent computational complexity. To address this challenge, we believe that an ideal candidate would be a machine model that grows linearly in the number of marker types included in the model. This makes such a task applicable to cancer prognosis, where many different markers are recorded over time. Generative models can account for disease trajectory shapes using components at the population, subpopulation, and individual levels, which simultaneously allows for heterogeneity across and within individuals and enables statistical strength to be shared across observations at different *resolutions* of the data [88]. Moreover, such systems can learn accurate and interpretable structured representations for disease trajectories by adapting their attention weights that determine the dependence of future disease states on past medical history.

Independent of the prediction model, the challenge is to continuously adapt and update in the face of progression heterogeneity. From the ML point of view, we recommend the use of a model capable of being applied dynamically in continuous time and updated as soon as any new data are available (eg, hidden Markov models). Such approaches can model the transition of disease stages or states, which implies that the progression is continuous, and the transition probability to the future state relies only on the current state and the time span. Instantiation of such causality-based ML was used to infer the underlying somatic staging of tumors from next-generation sequencing data [89].

Moving away from the modeling decision, the last challenge lies in the observed versus latent data artifacts. We believe that a powerful tool is an ML model that accounts for latent factors and covariates influencing disease expression, as standard regression models rely on observed features alone to explain variability [90]. For example, LSTM models over physiological word inputs from health records significantly improve performance as their representation encodes important information about what is *normal* for each physiological value or is more robust to sparseness in the physiological data.

Technological Interventions in Cancer Rehabilitation

Over the past decades, early diagnosis, new drugs, and more personalized treatment have led to impressive increases in survival rates of patients with cancer. However, the most mitigating side effects of commonly used therapies are a severe problem in oncology, leading to dose reduction, treatment delay, or discontinuation [91,92].

Opportunities

With the increasing number of cancer survivors, more attention is being paid to persistent sequelae of tumor treatment and supportive measures [93-95] used as adjuncts to mainstream cancer care to control symptoms and enhance well-being [96-100]. The broad literature overview allowed us to identify a series of challenges and the associated opportunities that digital interventions could offer in supportive care.

The first challenge we identified is the identification of therapeutic sequelae. This challenge offers the opportunity to develop interventions capable of assessing what deficits (sensory, motor, and/or cognitive) a specific patient has as a consequence of cancer therapy.

The next challenge arises when the intervention needs to quantify the magnitude of therapy sequelae. Here, we identify a clear opportunity to measure the level of deficit or dysfunction induced by the therapy. This is crucial in (1) designing the follow-up therapy scheme, (2) choosing a rehabilitation strategy, and (3) determining the therapy trajectory and dosage.

The last challenge we identified is the adaptive parametrization of rehabilitation. This challenge brings a valuable opportunity to take steps toward personalized treatment, namely, to parameterize the rehabilitation scheme according to the specific deficit type and level to drive rehabilitation.

Recommendations

To cope with patient sensory, motor, and cognitive deficit variability, it is necessary to perform a precise assessment of the 3 different dimensions. We believe that the 3 main challenges we identified as high-potential opportunities for digital interventions are also good candidates for ML algorithms. This technology can learn underlying correlations in patient data and generalize for robust prediction [101]. The mapping from the challenge to the recommendation is provided in the remainder of this section.

When tackling the identification of therapy sequelae, we recommend exploiting and mining large sets of structured and unstructured data describing a patient to identify correlations among various data types and how they map to a certain type of dysfunction. We propose the use of semisupervised techniques (eg, transductive support-vector machines [102]) that will only require a limited number of labels to generalize well.

Moreover, to address the magnitude of the therapy sequelae, we recommend the use of deep learning models, especially convolutional neural networks. This is because such networks are capable of learning relevant feature representation from unstructured data, such as pathological images or medical records. This, subsequently, allows deep learning methods to achieve good results in tasks such as regression, detection, and segmentation, which underlie the magnitude estimation.

Finally, addressing the challenge of adaptive parametrization of rehabilitation, we recommend using methods that combine learning capabilities for regressing arbitrary nonlinear functions (ie, deep learning—encoding the type of deficit covariance with the magnitude) and adaptation through guided searches in parameter spaces (ie, reinforcement learning—finding the best parameters of the rehabilitation scheme—dosage, type, and length that best fit the regressed function).

Addressing Current Interoperability Challenges

Patient data needed for the provision of the best treatment are often scattered across different systems rather than stored in a single location. Technologies that facilitate care coordination through interoperability are improving, but a seamless flow of information from one care setting to another still requires more progress. Without having a full picture, it is difficult to provide the best care in an era where cancer is considered to be a chronic disease, and there is an increased demand for consistent follow-up in terms of monitoring and early management of symptoms that indicate that cancer might have returned.

Interoperability is a primary consideration to achieve communication among applications, medical devices, and health care providers [103], although the growing demand for secondary use of clinical and administrative data increases the pressure to solve relevant challenges [104].

The currently popular use of EHRs has alleviated some of the barriers in using data from medical records for research, although fully interoperable electronic medical record systems are not yet a reality. Several efforts to develop and apply standards in the collection, extraction, and integration of data by standardization bodies, governments, the research community, and industry are in progress [105] with the aim to establish and adopt clinically relevant, integrated standards covering the entire oncology sector. Organizations such as Health Level Seven International [106] and Personal Connected Health Alliance Continua [107] help to deliver standards-based, open specifications that can support the flow of data from the point of capture into EHRs in the same format and coded content. Cancer-specific apps that exploit platforms and interoperability standards, such as SMART [108] and Fast Healthcare Interoperability Resources [109], are emerging [110-112].

Standards, such as those developed in the United States and Canada, to guide EHR vendors and public health central cancer registries in the implementation of standardized electronic reporting [113-115] can be used with third-party terminologies.

It is a fact that some data resolution may be lost during the process of mapping EHR fields to a formally described abstraction layer, which may be alleviated by the use of knowledge models such as ontologies as a knowledge background mapper; however, these common interfaces support queries across EHRs or the extraction of patient data in the same format to allow the merging of patient sets between numerous institutions.

Opportunities

Health information technology (IT) brings clinical data and patient information together and guides oncologists in making evidence-based care decisions that lead to improved outcomes. The potential benefits of interoperable interconnected tools and health systems are particularly important for oncology [116], as providing cancer-wide care depends on access to accurate and complete information as well as extensive coordination among patients, caregivers, and diverse provider groups through treatment and survivorship. Connecting the EHR has the potential to support diagnosis assistance for complex patients [117].

Data must be able to provide a complete look at the patient's medical history so that physicians can see what cancer medicines and treatments did or did not work. Clinicians also need to be able to avoid recommending the same procedure twice, prescribing a medicine the patient already tried, or missing results from a diagnostic test.

The need for consolidation and standardization efforts to create interoperable solutions [105] and the need for the cancer informatics community for national initiatives for data standardization and large-scale multidisciplinary research collaborations are timely and critical. In addition to supporting cancer care, cancer-related standards will help improve surveillance and research. Similar to migration from paper to electronic records, the shift toward data interoperability between EHR vendors may require policy changes [118,119].

Recommendations

As already proposed by [120], enabling interoperability among institutions and individuals that support care delivery across the cancer continuum is considered essential. Doing so requires developing, testing, disseminating, and adopting technical standards for information related to cancer care across the continuum to optimize the flow of information to serve the needs of caregivers, patients, and providers. To achieve this, standard open application programming interface platforms should be developed and used to facilitate the development of cancer-related apps.

As mentioned in [121], standards and protocols that aim to enhance the interoperability of different data sets are a highly relevant field for policy action. Incentives for the promotion of standards for the interoperability of clinical data, such as EHR or genome experiment data, should be developed, promoted, and incentivized to allow for the pooling of data, and comparison of system-level research should be provided for the comparison of research outcomes.

The open use and sharing of big data, without compromising patients' rights to privacy and confidentiality, should be promoted.

Patient-Clinician Shared Decision-Making Processes

Current clinical care is gradually moving toward patient-clinician shared decision making, as patient involvement can provide insights into best health states or outcomes in each case, apart from establishing a partnership that will help clinicians understand their patients' preferences [122]. In a complex disease, such as cancer, making a clinical decision is a difficult task because of the trade-off between the level of treatment benefits and the impact that adverse events or symptoms could have on the patient's life. This is especially crucial in cases where the expected medical outcomes are similar for different clinical procedures (ie, also referred to as *equipoise*), requiring an individualized and personalized health care process involving the clinician and the patient in a close interaction for the best decision making [123]. Nevertheless, actual CPGs still do not include or barely mention the impact of patient preferences along with clinical evidence.

Opportunities

It is necessary to overcome many barriers to ensure success in the inclusion of patient preferences during the decision-making process and to track their impact in the care services provided [124]. First, patient preferences should be considered as population knowledge that follows some general trends and not just as individual one-off cases or subjective and variable factors. This would allow including them along with the clinical evidence reported in the studies to identify *preference-sensitive* decisions (eg, decisions having lifelong implications or an uncertain benefit to the patient, unclear or conflicting evidence, risk of having side effects that negatively affect the patient's quality of life) of high levels of uncertainty about the best clinical procedure to follow [125].

Recommendations

As a complex disease, cancer involves many clinical specialties, with the need of creating a clear taxonomy (ie, systematic categorization) for patients' preferences that will serve as a standardization over all involved disciplines (eg, analysts, clinical psychologists). This kind of approach would help in the harmonization of the different points of view on the measurement of patients' preferences by labeling and extracting this information in a processable and understandable way [126,127]. Finally, with regard to the previous points, building a methodology to synthesize the current knowledge on preference trends to be able to describe preference-based decisions along with clinical-based evidence would have a great impact on health care performance, as it has been proven that preferences strongly influence the decision-making process [128,129].

Assessment of Clinical Evidence-Based Recommendations, Including PROMs

As presented in the previous sections, the evidence used in medical practice is based on clinical guidelines, which are often used as the evidence basis for the clinical DSS (CDSS) [130]. These guidelines are developed by multidisciplinary groups and are based on a systematic review of the scientific evidence, and their recommendations are explicitly linked to the supporting evidence and graded according to the strength of that evidence [131]. As reported by Harbor et al [131], the levels of evidence of the recommendations are based on study design and the methodological quality of individual studies, which is called the quality of evidence. Therefore, the latest approach also considers the strength of the recommendations, which is a trade-off between benefits and harms of a treatment, considering the quality of evidence [131,132]. Several oncology guidelines, such as the NCCN CPGs in Oncology [133] applied in DESIREE, provide quality of evidence measure, and it is up to the clinician to consider the strength of the recommendation depending on the specific patient situation. However, the subjective measures of patients with cancer, such as fatigue affected by oncological treatments, have not been systematically considered during the decision-making process. Making this information available during the decision-making process also provides other relevant and real-world quality measures to the treatment recommendation to best support clinicians and patients in their treatment decisions.

Opportunities

CDSSs provide the opportunity not only to quickly access the latest available evidence but also to incorporate new sources of information that can support the decision-making process. At the same time, the new IT era enables the acquisition of PROMs using questionnaires or even more sophisticated wearables that can measure activity, sleep, or other vital signs that can be translated to patient outcomes. These automatic ubiquitous technologies increase the knowledge required by patients. Processing these data and deducing the desired results will be very helpful. For example, tracking the daily activities of a user provides estimates of how active a patient is, which can be correlated with their depression and/or fatigue levels (variables that are gathered using questionnaires such as the European Platform for Cancer Research—Quality of Life Questionnaire with 30 items). Other techniques could also aid the acquisition of patient outcome information, such as the Ecological Momentary Assessment tests (short questionnaires that are sent to the patient frequently to obtain updates on their status). Thus, the collected data could be used to assess how good the treatment given was for each patient, not in the scope of a randomized control trial but in the real-world environment, without having a specific patient population, but the whole population.

Recommendations

In this context, in the DESIREE project focusing on primary breast cancer, new ways of including PROMs to assess guideline recommendations were explored. The aim of including PROMs does not conflict with the quality of evidence and strength of recommendation measurements but provides other quality attributes that contribute to the decision-making process, considering the patient status reported by the patients themselves.

There are some specifications that should be considered for this quality assessment, which could be named questionnaire based–patient-reported outcome measures, and they include (1) consideration of standard questionnaires for each medical case, such as the International Consortium for Health Outcomes Measurement questionnaires [134], (2) definition of specific protocols to assess the results of the questionnaires, and (3) determination of how these results should be used during the decision-making process.

Ambiguity on Clinical Guidelines Used for Clinical Decision Support

When implementing CPGs, several characteristics must be considered to ensure both good health care quality levels and clinicians' satisfaction. They must assure the validity and reliability of their clinical content, along with their clinical applicability in real clinical settings, and must be clear when defining the procedures to be followed in the current clinical performance procedures within a health care system [135,136]. Nevertheless, several barriers cause the dissemination of the guidelines to be tedious and difficult, mainly because of the ambiguity of the knowledge defined in them. The lack of awareness and familiarity with the recommendations provided in the guideline, a lack of agreement due to different clinical interpretations and simplification of the clinical knowledge reported in the guidelines, and the lack of outcome expectancy are some of the reported barriers that cause a lower adherence of clinicians to guidelines and have an effect on their implementation, compliance, and adherence in real clinical settings [137,138].

Opportunities

Actual trends move toward highly interactive computerized systems focusing on intuitively presenting complex clinical cases, where clinicians may access and check computerized clinical data and take away insights from all of this information in a more natural and intuitive way, alleviating the ambiguities of the guidelines through a data-driven approach guided by previous practice [139]. The digital implementation of the CPGs provides evidence-based decision support (ie, computer-interpretable guidelines [CIGs]) [140]. To achieve this, the knowledge available has to be formalized in a manner that is correct and of good quality, by following a consistent and adequate methodological workflow of the clinical processes and objectives, trying to reduce or provide some solutions to the paper-based guidelines ambiguities [141].

Recommendations

The proposed directions for realizing these needs include the promotion of standardized clinical terminology that facilitates the understanding and univocal interpretation of the clinical data to be analyzed and the clinical knowledge formalized in the CIGs (eg, the Breast Imaging Reporting and Database System standard for breast anomalies [142]); to be effective, clinical guidelines need to be integrated with the care flow and should provide patient-specific advice when and where needed. Hence, their formalization as CIGs should make it possible to develop CIG-based DSSs, which have a better chance of affecting clinician behavior than narrative guidelines. This will help in providing optimal personalized guideline-based recommendations, avoiding ambiguities, and at a reasonable cost and implementation effort [143].

Up-to-Date Clinical Evidence Guidelines for CDSS

Clinical guidelines are tedious to develop, and it is even more difficult to interpret and put them into a computer-interpretable way. This usually requires the close collaboration of knowledge engineers and medical experts. A closely related and important issue is the guideline development process or how CPG development working groups are composed. Usually, these teams comprise quality auditors or managers who are guided by their opinions, interests, and experience and intend to formalize evidence, seeking appropriateness of the provided recommendations but ignore the iterative and causal reasoning of clinicians [144]. Depending on the clinical context and according to the approaches followed for developing, disseminating, and implementing them in practice, CPGs can be more or less successful when reporting the latest clinical evidence [145]. Thus, the rapid advances in clinical practice have made the task of updating the guidelines used for the CDSS more challenging. In particular, in oncology, the number of discoveries is increasing rapidly. As reported by Beatty [146], drug label information and indications do not always keep up with meaningful advances in oncology resources. However, clinicians may be aware of the potential benefits of a particular therapy. Hence, Beatty [146] reports that continuous vigilance (in the form of continuing medical education and literature review) is a survival attribute for medical oncologists. Consequently, there is also a challenge to update the CIGs and have up-to-date guideline-based CDSS.

Opportunities

In this context, it is crucial to have tools supporting the easy updating of CIGs used for CDSS by clinicians themselves or knowledge engineers who do not necessarily have to understand the technology and/or the programming language used. Thus, interfaces that are easy to use and easy to understand are required for this purpose. This limitation was highlighted in the DESIREE project; however, this is an issue not only for cancer care but also for other clinical specialties.

Recommendations

Therefore, we propose an authoring tool for CPG formalization [147] that should at least fulfill the following requirements: (1) enable the input of guidelines' information in an easy and visual manner, such as the form of rules or flowcharts, and (2) enable the modification of CIGs previously formalized in the system. In addition, future work should focus on providing a tool for detecting modifications in guidelines as well as semiautomate the formalization of guidelines, using natural language processing.

Trust and Reliance on Cancer Care

The change in both hardware and software over the past decade has been remarkable. Equally noteworthy has been the ever-increasing internet speed and the accompanying growth in the demand for connectivity. This growth and development have increased our ability to take on challenging projects to improve early diagnosis and improve the quality of cancer care. However, most discussions relating to health data and associated analytical tools often emphasize data privacy and security at the expense of other topics. These talks often overlook the dynamic nature of both health data and the software used in the analysis of those data. In addition, the popularity of internet-based apps and the use of such tools by patients for self-diagnosis necessitate a call to action. There is a need to examine the reliability and trustworthiness of health-related computational tools used in diagnosis and DSS by health care providers. Here, we will focus on the reliance and reliability of computational tools with an emphasis on cancer care and diagnosis.

Opportunities

Reliance and reliability of information in health care are incredibly important. These attributes are particular when it comes to genetic information associated with cancer, which is highly critical in the development of optimal clinical intervention strategies. With an increasing number of people falling victims to medical misinformation and propaganda on the internet [148], advocacy to develop reporting and assessment standards has been long overdue. The internet undoubtedly provides patients with countless sources of health information related to cancer diagnosis and care, some of which represent the gold standard, whereas other sources remain to be of ill repute. The internet promotes decentralization and web democratization of access. Having an open-source or access model certainly helps to promote this agenda. These models provide help and are available to many without the control of smaller powerful broker or agents. However, it challenges the core process of software development, which includes specification, design, development, verification, validation, and management. Although inspection or peer review is the method of choice to check for static processes and software testing a true validation and verification method to check the system, the former fails to check emergent properties such as reliability and performance.

There is no turning back from this path of dependence on the tools and information available via the internet. People have a reasonable expectation of establishing trust and validation of these tools. Initiatives such as the Secure Socket Layer certification system, implemented for encrypting sensitive information sent across the internet or the use of digital badge to indicate or attest to adhere to an acceptable standard and/or individual skill competency (*seal of approval* from prominent institutions) or the use of the message-digest algorithm 5 (MD5) hash to ensure data integrity work to build and maintain that trust. Health IT requires a method to assure the credibility of the results generated by various computational tools available on the web. Unfortunately, none of the previously listed solutions are applicable to health software because of the dynamic nature of data as well as the software (as many deviate from the original specifications for which acceptable test and peer review of the result exist).

Recommendations

Having reliable information alone is not sufficient for people to construct their foundation of trust. People may have a distrust on the reliability of official information on certain topics because of previous unpleasant experiences. One conclusion that can be drawn from our experience in developing relevant computational tools over the past decade is the questionable credibility of the results obtained from such tools. A review of the literature shows a great deal of volatility in the availability of health IT resources and shows that providing explanations for software errors is an acceptable approach. Building on this notion, a certificate that discloses who is responsible and what tests are done or can be done to validate or test the trustworthiness of the output, something along the line of the MD5 checksum, could be envisioned as an acceptable solution to address this issue. Such a certificate could include the versions of the data and the software in a report to help explain the deviation from the previous version. This could be seen as a reasonable step in the right direction before the availability of peer-approved permanent solutions.

Trust in Computer-Aided Diagnosis Systems

An important issue emerging for decision support in medical diagnosis is the trust that clinicians might have in the outputs generated by a computer-aided diagnostic system. This is a highly relevant issue for tissue characterization in general and image-based tumor classification in particular. Hence, it should be considered whether the metrics such as accuracy (ACC) and area under the curve (AUC) correlate well with confidence in the algorithms used in computer-aided design (CAD) systems. If the CAD system provides a recommendation with moderate or low confidence, then a radiologist may deem the recommendation to be useless, even if the classifier being used has a high overall value for ACC and/or AUC. However, if the confidence is much higher, the clinician may deem the recommendation more useful in supporting their decision making. Therefore, building CAD systems in which clinicians have confidence is essential if those CAD systems are to be adopted and are to play a role in fully exploiting the range of digital information available for assisting diagnosis.

Recent studies discuss the failure of CAD systems in terms of a lack of *trust* by clinicians in the outputs that the CAD systems generate [149-151]. When designing and evaluating an algorithm within a CAD system, technical developers tend to report metrics such as ACC and AUC to demonstrate the performance of a classification method. However, such metrics do not measure the degree of confidence in individual recommendations made by the CAD system. For example, a CAD system may produce a high AUC value very close to the optimal value of 1, but most of the cases might be classified with low confidence (as measured by the likelihood or probability value associated with the recommended class). Indeed, studies have shown that the use of CAD systems that incorporate inadequate metrics can be detrimental to diagnosis [152].

Opportunities

Exploratory studies for breast mass classification [153] and classification of microcalcifications in mammograms [154] using data sets taken from the Curated Breast Imaging Subset of DDSM [155] showed that although most classifiers produce similar overall ACC and AUC values, their performances differ significantly in terms of confidence measure. High ACC or AUC does not provide a full indication of the confidence level of a CAD system. This aspect of the usability of CAD methods is a key but overlooked challenge. Most cancer CAD systems have used ACC and AUC as the main evaluation metrics; however, the probability outputs of the classifier must also be considered and harnessed to measure the degree of certainty of the system in its decision making. Hence, if clinicians are to have confidence in the support provided by a CAD system, the classifier that is chosen to be implemented as part of the system should not necessarily be the one that performs best overall using the standard metrics. CAD support systems must embody reliable confidence measures as one of their key elements [156]. It is therefore essential that the domain of trust in CAD systems should be explored to incorporate trust into the initial classifier design when such algorithms are to be embedded into a cancer CAD system.

Recommendations

To fully exploit the range of digital information available for assisting diagnosis, it is important to identify and implement specific actions to increase the trust of physicians in cancer CAD systems and overcome the barriers to adoption of such systems. Most cancer CAD systems have used ACC and AUC as the main performance evaluation metrics; however, confidence measures must also be considered, as the traditionally used metrics are inadequate for informing clinicians in terms of the confidence that they might place on the recommendations provided. Besides, research in the design of classifiers that are incorporated into CAD systems is essential if future CAD systems are to be trusted by clinicians and adopted as a valued and reliable, and indeed routine, element of the cancer diagnosis process.

Regulatory Roadmap for Validating the Effectiveness of AI–Based Models for Clinical Decision Making

Data and image analysis algorithms usually require regulatory approval by the Food and Drug Administration (FDA) in the United States and conformity assessment leading to a Conformité Européenne mark in the EU [157], under the new Medical Device Regulation. Both EU and US regulations classify medical devices according to risk classes, and those that fall under certain risk classes are required to undergo clinical trials to be marketed [158]. To date, these guidelines target systems with so called *locked* algorithms, that is, algorithms whose functions do not change, rather than *adaptive* ones, that is, whose behavior can change over time based on new data [159]. Adaptive AI–based tools have the potential to autonomously “adapt and optimize device performance in real-time to continuously improve healthcare for patients.” The FDA has already approved several AI- or ML-based software as a medical device (SaMD), but these have only included algorithms that are *locked* before marketing [160].

Opportunities

Regulations that aim to validate AI in a safe and transparent way must consider doing so without compromising the potential of AI. Moreover, characteristics of AI that may be seen as risks such as biases in data, its *blackbox nature*, and model degradation, among others, have to be included in the validation process to better ensure safety and build public trust [161]. The FDA recently proposed an approach that aimed to “have tailored, pragmatic, and least burdensome regulatory oversight” while validating the continued safety, effectiveness, and performance of SaMDs. The framework recommends (1) establishing clear expectations of quality systems and good ML practices, for example, demonstrating analytical and clinical validation, (2) creating SaMD prespecification sheets and algorithm change protocols to qualify for a premarket review on the safety and effectiveness of the AI-based tool, (3) preparing an approach for *AI- or ML-based software modifications* (eg, data for retraining), and (4) implementing mechanisms that support *transparency and real-world performance monitoring*.

In the EU, there are no regulatory guidelines that specifically cover AI in health care. Nevertheless, draft ethics guidelines have been published by the High-Level Expert Group on AI in April 2019, proposing 7 key requirements that AI systems should meet to be realized as *trustworthy AI*: human agency and oversight, including fundamental rights; technical robustness and safety; privacy and data governance; transparency, diversity, nondiscrimination, and fairness [162,163]; societal and environmental well-being; and accountability. The guidelines also propose technical and nontechnical methods to implement the 7 key requirements and recommend that these should be continuously evaluated and addressed throughout the AI system’s life cycle.

Recommendations

The frameworks enumerated above are, however, still being piloted or discussed. To properly define a regulatory framework for users, stakeholders and use cases (data flows) should be identified and defined. New regulatory frameworks should be built to provide guidance for the validation or qualification of AI tools within different scenarios and pathways (for nonclinical or preclinical use or clinical use), taking into consideration the adaptive nature of AI-based tools. The framework should consolidate input from scientific experts and health authorities and should take into consideration published relevant guidelines, for example, from the High-Level Expert Group on AI, the Medical Device Coordination Group and relevant implementing EU legislation, and international regulations (eg, from the FDA).

Table 1 summarizes the various topics discussed, the related works for each topic, the opportunities, and the recommendations from the workshops.

Table 1. Topics, opportunities, and recommendation in cancer care.

Topic or section (references)	Opportunities	Recommendations
Patient engagement and participatory design [11-29]	• Involvement of real users • Identification of user needs • Unique perspective on user acceptability, usability, and feasibility	• Participatory design approach early and throughout the design process • Focus groups with stakeholder representatives • Fuse findings with those from other sources
Small data analytics [36]	• Address the annotation problem via appropriate tools • Enable experts to teach ML[a] models that automatically build and annotate their data sets • Automatically represent knowledge in a structured and computerized way	• Design new machine learning algorithms that needs minimal feedback from human experts • Use knowledge-based learning that can be extended by data-driven findings easily and that uses standardized terminologies to provide interoperability and ease the updating and maintenance of the latest evidence
Integration and data management [42-50]	• Exploiting aggregated, heterogeneous, and distributed data • Translating the clinical findings into an intuitive representation for the patient • Building individualized clinical recommendations based on the data interpretation projected in actions upon the patient • Data completeness and augmentation	• Embed continuous user feedback and iterative prototyping in the intervention • Usage of tools to extract and represent the medical substrate by synthesizing only relevant aspects in a declarative way • Development of clinical projections from individualized patient recommendations to therapy plans that embed temporal, procedural, and reasoning processes • Incorporation of lived experiences of the patients
Extracting patient's portrait [62-67]	• Exploit correlations among multiple data sources to extract patient profile • Use data mining and machine learning to guide therapeutic schemes • Identify relevant genetic, phenotypical, physiological, lifestyle, and medical data correlations with diagnosis • Provide an integrative approach to patient-centered data and demonstrate the potential of feature selection in data analysis and predictive patient-specific outcomes	• Exploit available data • Extract underlying correlations • Integrate the multitude of representations in a structured object guiding therapeutic interventions in cancer care
Learning patient disease trajectory for personalized diagnosis [81-84]	• Handling data and choosing specific latent variable models to summarize and extract information from the irregularly sampled and sparse data • Learning of a disease trajectory is linked to the inherent computational complexity • Continuous adaptation and update in face of disease progression heterogeneity • Observed versus latent data artifacts	• Use of discriminative models that exploit conditions on marker histories instead of jointly modeling them • Focus on machine models which grow linearly in the number of marker types included in those models • Use of a model capable of being applied dynamically in continuous time and updated • Exploit models that account for latent factors and covariates influencing disease expression
Technological interventions in cancer rehabilitation [91-102]	• Cope with patient sensory, motor, and cognitive deficit variability • Identify therapy sequelae	• Perform a precise assessment of patient's sensory or motor or cognitive deficit variability • Use machine learning algorithms to identify underlying correlations in patient data and generalize for robust prediction • Exploit and mine large sets of structured and unstructured data to identify correlations and map to a certain type of dysfunction
Addressing current interoperability challenges [103-111,113-115]	• Provide cancer-wide care • Support diagnosis assistance for complex patients • Provide a complete look at the patient's medical history so physicians can see ineffective treatments • Improve surveillance and research	• Develop, test, disseminate, and adopt technical standards for information related to cancer care across the continuum • Optimize the flow of information to serve the needs of caregivers, patients, and providers • Develop and use standard, open application programming interfaces • Promote incentives for the pooling of data and comparison of system-level research • Support open use and sharing of big data, without compromising patients' rights to privacy and confidentiality

Topic or section (references)	Opportunities	Recommendations
Patient-clinician shared decision-making processes [122-129]	• Inclusion of patient preferences during the decision making and tracking of their impact in the provided care services • Identification of *preference-sensitive* decisions	• Create a clear taxonomy (ie, systematic categorization) for patients' preferences to serve as a standardization • Harmonize different points of view to facilitate labeling and extraction of information in a processable and understandable way • Build a methodology to synthesize knowledge
Assessment of clinical evidence-based recommendations, including PROMs[b] [130-133]	• Quick access to latest available evidence • Incorporate new sources of information that can support the decision-making process • Increase the knowledge required by patients • Effectively assess how good the treatment given was for each patient, not in the scope of a randomized control trial, but in the real-world environment	• Explore new ways of including PROMs to assess guideline recommendations • Exploit PROMs in the decision-making process, considering patient status reported by the patients themselves • Consider and use existing quality assessment specifications for PROMs
Ambiguity on clinical guidelines used for clinical decision support [135-138]	• Insight from complex clinical cases in a natural and intuitive way • Patient-specific advice when and where needed	• Promotion of standardized clinical terminology • Integration of clinical guidelines with care flow
Up-to-date clinical evidence guidelines for CDSS[c] [144-146]	• Create tools that support the easy updating of CIGs[d] for clinicians • Interfaces that are easy to use and understand are required for this purpose for CIGs	• Generate tools that enable the input of guideline information in an easy and visual manner and enable the modification of CIGs previously formalized in the system • Provide a tool for detecting modifications on guidelines • Semiautomate the formalization of guidelines using natural language processing
Trust and reliance on cancer care [148]	• Assurance of the credibility of results generated by various computational tools available on the web	• Provision of a certificate that discloses who is responsible and what tests are done or can be done to validate or test the trustworthiness of the output • Include the versions of the data and the software in a report to help explain the deviation from the previous version
Trust in computer-aided diagnosis systems [149-152]	• Increase confidence in the support provided by a CAD[e] system	• CAD support systems must embody reliable confidence measures as one of their key elements • Incorporate trust into the initial classifier design when such algorithms are to be embedded into a cancer CAD system
Regulatory roadmap for validating the effectiveness of AI[f]-based models for clinical decision making [157-160]	• Validation of AI in a safe and transparent way without compromising the potential of AI	• Identify and define users, stakeholders, and use cases (data flows) • Build regulatory frameworks aiming to provide guidance toward the validation or qualification of AI tools within different scenarios and pathways • Consolidate input from scientific experts, health authorities, and published guidelines

[a] ML: machine learning.

[b] PROM: patient-reported outcome.

[c] CDSS: clinical decision support system.

[d] CIG: computer-interpretable guideline.

[e] CAD: computer-aided design.

[f] AI: artificial intelligence.

Discussion

This paper presents the key topics that were discussed as part of two international workshops on the current status of technological and data-driven innovations in cancer care. Key challenges and opportunities have been identified, and several recommendations have been made to facilitate the acceleration of progress in the data-driven management of cancer. The workshops presented the work that was being conducted in four Horizon 2020 EU–funded projects, namely BOUNCE, CATCH ITN, DESIREE, and MyPal. These projects provide a rich landscape of the challenges and opportunities of the current state of the art of new technologies in cancer care. The authors have presented these issues and discussed recommendations that can be used for further research as well as practical implementation of such tools in cancer.

Acknowledgments

The work presented in this paper is part of the BOUNCE project, which has received funding from the EU's H2020 Research and Innovation Program under grant agreement number 777167.

Conflicts of Interest

None declared.

References

1. Lin L, Yan L, Liu Y, Yuan F, Li H, Ni J. Incidence and death in 29 cancer groups in 2017 and trend analysis from 1990 to 2017 from the Global Burden of Disease Study. J Hematol Oncol 2019 Sep 12;12(1):96 [FREE Full text] [doi: 10.1186/s13045-019-0783-9] [Medline: 31511035]
2. Technological and data-driven innovations in cancer care. tech4Cancer. URL: https://www.tech4cancer.com/ [accessed 2020-11-21]
3. Technological and Data-Driven Innovations in Cancer Care. CBMS 2019. 2020 Nov 21. URL: https://www.uco.es/congresos/cbms2019/tech4cancer/ [accessed 2020-11-24]
4. 2020 Nov 21. URL: https://bibe2019.ics.forth.gr/participation/technological-and-data-driven-innovations-in-cancer-care/ [accessed 2020-11-21]
5. Kondylakis H. Developing a Data Infrastructure for Enabling Breast Cancer Women to BOUNCE Back. 2019 Jun 5 Presented at: IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS); June 5, 2019; Cordoba, Spain. [doi: 10.1109/cbms.2019.00134]
6. Kourou K. Computational Models for Predicting Resilience Levels of Women with Breast Cancer. In: Mediterranean Conference on Medical and Biological Engineering and Computing. 2019 Sep 26 Presented at: XV Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON); 2019; Coimbra, Portugal. [doi: 10.1007/978-3-030-31635-8]
7. Desiree-project. URL: http://www.desiree-project.eu/ [accessed 2020-11-21]
8. MyPal. URL: https://mypal-project.eu/ [accessed 2020-11-21]
9. Maramis C, Karamanidou C, Schera F. Using Electronic Patient Reported Outcomes to Foster Palliative Cancer Care: the MyPal approach. 2020 Oct 28 Presented at: IEEE International Conference on Bioinformatics and BioEngineering; Oct 28-30, 2019; Athens, Greece. [doi: 10.1109/bibe.2019.00079]
10. CATCH-ITN. URL: https://www.catchitn.eu/ [accessed 2020-11-21]
11. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
12. Kondylakis H, Koumakis L, Tsiknakis M, Kiefer S. Personally managed health data: barriers, approaches and a roadmap for the future. J Biomed Inform 2020 Jun;106:103440. [doi: 10.1016/j.jbi.2020.103440] [Medline: 32445857]
13. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002 Feb;4(1):e3 [FREE Full text] [doi: 10.2196/jmir.4.1.e3] [Medline: 11956035]
14. Karekla M, Kasinopoulos O, Neto DD, Ebert DD, Van Daele T, Nordgreen T, et al. Best practices and recommendations for digital interventions to improve engagement and adherence in chronic illness sufferers. Eur Psychol 2019 Jan;24(1):49-67. [doi: 10.1027/1016-9040/a000349]
15. Canada's Strategy for Patient-Oriented Research. Canadian Institute of Health Research. 2011. URL: https://cihr-irsc.gc.ca/e/44000.html [accessed 2020-11-21]
16. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
17. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2016 Dec 13:254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
18. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen J. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

19. Sieverink F, Kelders SM, van GJE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
20. Craig LR, Tada Y, Hilfiker S, Baur C. The assessment of user engagement with eHealth content: the eHealth engagement scale. J Comput Mediat Commun 2010 Jun 01;15(4):681. [doi: 10.1111/j.1083-6101.2009.01514.x]
21. Dewar AR, Bull TP, Malvey DM, Szalma JL. Developing a measure of engagement with telehealth systems: the mHealth Technology Engagement Index. J Telemed Telecare 2017 Feb;23(2):248-255. [doi: 10.1177/1357633X16640958] [Medline: 27036878]
22. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
23. Phillips JL, Currow DC. Cancer as a chronic disease. Collegian 2010;17(2):47-50. [doi: 10.1016/j.colegn.2010.04.007] [Medline: 20738055]
24. Menichetti J, Graffigna G. 'PHE in Action': development and modeling of an intervention to improve patient engagement among older adults. Front Psychol 2016;7:1405 [FREE Full text] [doi: 10.3389/fpsyg.2016.01405] [Medline: 27695435]
25. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015 Jan;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
26. Lauffenburger JC, Lewey J, Jan S, Nanchanatt G, Makanji S, Ferro CA, et al. Rationale and design of the ENhancing outcomes through goal assessment and generating engagement in diabetes mellitus (ENGAGE-DM) pragmatic trial. Contemp Clin Trials 2017 Aug;59:57-63. [doi: 10.1016/j.cct.2017.05.014] [Medline: 28571997]
27. Rise MB, Eriksen L, Grimstad H, Steinsbekk A. The long-term effect on mental health symptoms and patient activation of using patient feedback scales in mental health out-patient treatment. A randomised controlled trial. Patient Educ Couns 2016 Jan;99(1):164-168. [doi: 10.1016/j.pec.2015.07.016]
28. Fergusson D, Monfaredi Z, Pussegoda K, Garritty C, Lyddiatt A, Shea B, et al. The prevalence of patient engagement in published trials: a systematic review. Res Involv Engagem 2018 May 22;4:17. [doi: 10.1186/s40900-018-0099-x]
29. Karamanidou C, Natsiavas P, Koumakis L, Marias K, Schera F, Schäfer M, et al. Electronic patient-reported outcome-based interventions for palliative cancer care: a systematic and mapping review. JCO Clin Cancer Inform 2020 Jul;4:647-656 [FREE Full text] [doi: 10.1200/CCI.20.00015] [Medline: 32697604]
30. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol 2011 Mar;8(3):184-187. [doi: 10.1038/nrclinonc.2010.227] [Medline: 21364692]
31. Kondylakis H, Koumakis L, Mazzocco K, Tsiknakis M, Marias K. Participatory aspects of ICT infrastructures for cancer management. In: P5 eHealth: An Agenda for the Health Technologies of the Future. Cham: Springer; Nov 30, 2019:A.
32. Nieroda ME, Lophatananon A, McMillan B, Chen L, Hughes J, Daniels R, et al. Online decision support tool for personalized cancer symptom checking in the community (REACT): acceptability, feasibility, and usability study. JMIR Cancer 2018 Jul 04;4(2):e10073 [FREE Full text] [doi: 10.2196/10073] [Medline: 29973334]
33. Karamanidou C, Maramis C, Stamatopoulos K, Koutkias V. Development of a ePRO-based palliative care intervention for cancer patients: a participatory design approach. Stud Health Technol Inform 2020 Jun 16;270:941-945. [doi: 10.3233/SHTI200300] [Medline: 32570520]
34. Saunders R, Stuckhardt L. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington (DC): National Academies Press (US); 2013.
35. van Gemert-Pijnen J, Nijland N, van LM, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
36. Ehl C. Data — the Fuel for Artificial Intelligence. Medium. URL: https://medium.com/@cehl/data-the-fuel-for-artificial-intelligence-ed90bf141372 [accessed 2020-11-21]
37. LionBridge AI. URL: https://lionbridge.ai/ [accessed 2020-11-21]
38. openvinotoolkit / cvat. GitHub. URL: https://github.com/opencv/cvat [accessed 2020-11-21]
39. Amazon Mechanical Turk. URL: https://www.mturk.com/ [accessed 2020-11-21]
40. National Comprehensive Cancer Network. URL: https://www.nccn.org/professionals/physician_gls/default.aspx [accessed 2020-11-21]
41. ESMO Guidelines. ESMO. URL: https://www.esmo.org/guidelines [accessed 2020-11-21]
42. Dwamena F, Holmes-Rovner M, Gaulden C, Jorgenson S, Sadigh G, Sikorskii A. Interventions for providers to promote a patient-centred approach in clinical consultations. Cochrane Database Syst Rev 2012;12:CD003267. [doi: 10.1002/14651858.cd003267.pub2]
43. Wolpin S, Halpenny B, Whitman G, McReynolds J, Stewart M, Lober W, et al. Development and usability testing of a web-based cancer symptom and quality-of-life support intervention. Health Informatics J 2014 Jan 09;21(1):10-23. [doi: 10.1177/1460458213495744]

44. Klasnja P, Civan HA, Unruh K, Pratt W. Blowing in the wind: unanchored patient information work during cancer care. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Apr Presented at: CHI '10; April 2010; Atlanta p. 193-202. [doi: 10.1145/1753326.1753355]
45. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
46. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
47. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
48. Krist AH, Woolf SH. A vision for patient-centered health information systems. J Am Med Assoc 2011 Jan 19;305(3):300-301 [FREE Full text] [doi: 10.1001/jama.2010.2011] [Medline: 21245186]
49. Mulder I, Schikhof Y, Vastenburg M, Card A, Dunn T, Komninos A, et al. Designing with care: the future of pervasive healthcare. IEEE Pervasive Comput 2009 Oct;8(4):85-88. [doi: 10.1109/mprv.2009.71]
50. Bergeson SC, Dean JD. A systems approach to patient-centered care. J Am Med Assoc 2006 Dec 20;296(23):2848-2851. [doi: 10.1001/jama.296.23.2848] [Medline: 17179462]
51. Rance B, Canuel V, Countouris H, Laurent-Puig P, Burgun A. Integrating heterogeneous biomedical data for cancer research: the CARPEM infrastructure. Appl Clin Inform 2016;7(2):260-274 [FREE Full text] [doi: 10.4338/ACI-2015-09-RA-0125] [Medline: 27437039]
52. Wysham NG, Wolf SP, Samsa G, Abernethy AP, LeBlanc TW. Integration of electronic patient-reported outcomes into routine cancer care: an analysis of factors affecting data completeness. JCO Clin Cancer Inform 2017 Nov;1:1-10 [FREE Full text] [doi: 10.1200/CCI.16.00043] [Medline: 30657370]
53. König IR, Malley J, Pajevic S, Weimar C, Diener HC, Ziegler A. Patient-centered yes/no prognosis using learning machines. Int J Data Min Bioinform 2008;2(4):289-341 [FREE Full text] [doi: 10.1504/ijdmb.2008.022149] [Medline: 19216340]
54. Spector J, Merrill M, Elen J, Bishop M. Handbook of Research on Educational Communications and Technology: A Project of the Association for Educational Communications and Technology. United States: Springer; 2014.
55. Ruland CM, White T, Stevens M, Fanciullo G, Khilani SM. Effects of a computerized system to support shared decision making in symptom management of cancer patients: preliminary results. J Am Med Inform Assoc 2003;10(6):573-579 [FREE Full text] [doi: 10.1197/jamia.M1365] [Medline: 12925545]
56. Nicora G, Vitali F, Dagliati A, Geifman N, Bellazzi R. Integrated multi-omics analyses in oncology: a review of machine learning methods and tools. Front Oncol 2020;10:1030 [FREE Full text] [doi: 10.3389/fonc.2020.01030] [Medline: 32695678]
57. Zhao B. Clinical data extraction and normalization of cyrillic electronic health records via deep-learning natural language processing. JCO Clin Cancer Inform 2019 Sep;3:1-9 [FREE Full text] [doi: 10.1200/CCI.19.00057] [Medline: 31577448]
58. González-Ferrer A, ten Teije A, Fdez-Olivares J, Milian K. Automated generation of patient-tailored electronic care pathways by translating computer-interpretable guidelines into hierarchical task networks. Artif Intell Med 2013 Feb;57(2):91-109. [doi: 10.1016/j.artmed.2012.08.008]
59. Fdez-Olivares J, Cózar J, Castillo L. Oncotheraper: Clinical decision support for oncology therapy planning based on temporal hierarchical tasks networks. 2008 Presented at: Workshop on Knowledge Management for Health Care Procedures; 2008; Patras, Greece p. 25-41. [doi: 10.1007/978-3-642-03262-2_3]
60. Williams F, Jeanetta SC. Lived experiences of breast cancer survivors after diagnosis, treatment and beyond: qualitative study. Health Expect 2016 Jun;19(3):631-642. [doi: 10.1111/hex.12372] [Medline: 25953316]
61. Tanay MA, Armes J. Lived experiences and support needs of women who developed chemotherapy‐induced peripheral neuropathy following treatment for breast and ovarian cancer. Eur J Cancer Care 2019 Feb 20;28(3):e13011. [doi: 10.1111/ecc.13011]
62. Kirk R, Villanueva MT, Hutchinson L. Patient profiling in breast cancer—is knowledge power? Nat Rev Clin Oncol 2011 Dec 21;9(1):1-1. [doi: 10.1038/nrclinonc.2011.196]
63. Halpern Y, Horng S, Choi Y, Sontag D. Electronic medical record phenotyping using the anchor and learn framework. J Am Med Inform Assoc 2016 Jul;23(4):731-740 [FREE Full text] [doi: 10.1093/jamia/ocw011] [Medline: 27107443]
64. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. Br Med J 2015 Apr 24;350:h1885 [FREE Full text] [doi: 10.1136/bmj.h1885] [Medline: 25911572]
65. Goletsis Y, Exarchos T, Giannakeas N, Fotiadis D. Intelligent patient profiling for diagnosis, staging and treatment selection in colon cancer. 2008 Presented at: IEEE International Conference on BioInformatics and BioEngineering; October 8-10, 2008; Athens, Greece p. 1-6. [doi: 10.1109/bibe.2008.4696785]
66. Ng K, Sun J, Hu J, Wang F. Personalized predictive modeling and risk factor identification using patient similarity. AMIA Jt Summits Transl Sci Proc 2015;2015:132-136 [FREE Full text] [Medline: 26306255]

67. Elissen AM, Hertroijs DF, Schaper N, Vrijhoef HJ, Ruwaard D. Profiling patients' healthcare needs to support integrated, person-centered models for long-term disease management (profile): research design. Int J Integr Care 2016 Apr 29;16(2):1 [FREE Full text] [doi: 10.5334/ijic.2208] [Medline: 27616957]
68. Tuxen IV, Rohrberg KS, Oestrup O, Ahlborn LB, Schmidt AY, Spanggaard I, et al. Copenhagen prospective personalized oncology (CoPPO)-clinical utility of using molecular profiling to select patients to phase I trials. Clin Cancer Res 2019 Feb 15;25(4):1239-1247 [FREE Full text] [doi: 10.1158/1078-0432.CCR-18-1780] [Medline: 30274980]
69. Lahat D, Adali T, Jutten C. Multimodal data fusion: an overview of methods, challenges, and prospects. Proc IEEE 2015 Sep;103(9):1449-1477. [doi: 10.1109/jproc.2015.2460697]
70. Yuan Y, Savage RS, Markowetz F. Patient-specific data fusion defines prognostic cancer subtypes. PLoS Comput Biol 2011 Oct;7(10):e1002227 [FREE Full text] [doi: 10.1371/journal.pcbi.1002227] [Medline: 22028636]
71. Rubinstein WS, Acheson LS, O'Neill SM, Ruffin MT, Wang C, Beaumont JL, Family Healthware Impact Trial (FHITr) Group. Clinical utility of family history for cancer screening and referral in primary care: a report from the Family Healthware Impact Trial. Genet Med 2011 Nov;13(11):956-965 [FREE Full text] [doi: 10.1097/GIM.0b013e3182241d88] [Medline: 22075527]
72. Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinformatics 2017 Jan 3;18(9):1. [doi: 10.1186/s12859-016-1423-9]
73. El-Manzalawy Y, Hsieh T, Shivakumar M, Kim D, Honavar V. Min-redundancy and max-relevance multi-view feature selection for predicting ovarian cancer survival using multi-omics data. BMC Med Genomics 2018 Sep 14;11(Suppl 3):71 [FREE Full text] [doi: 10.1186/s12920-018-0388-0] [Medline: 30255801]
74. Zhang C, Biś D, Liu X, He Z. Biomedical word sense disambiguation with bidirectional long short-term memory and attention-based neural networks. BMC Bioinformatics 2019 Dec 02;20(Suppl 16):502 [FREE Full text] [doi: 10.1186/s12859-019-3079-8] [Medline: 31787096]
75. Beeksma M, Verberne S, van den Bosch A, Das E, Hendrickx I, Groenewoud S. Predicting life expectancy with a long short-term memory recurrent neural network using electronic medical records. BMC Med Inform Decis Mak 2019 Feb 28;19(1):36 [FREE Full text] [doi: 10.1186/s12911-019-0775-2] [Medline: 30819172]
76. Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: a deep learning approach. J Biomed Inform 2017 May;69:218-229 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.001] [Medline: 28410981]
77. Tranchevent L, Azuaje F, Rajapakse JC. A deep neural network approach to predicting clinical outcomes of neuroblastoma patients. BMC Med Genomics 2019 Dec 20;12(Suppl 8):178 [FREE Full text] [doi: 10.1186/s12920-019-0628-y] [Medline: 31856829]
78. Khosravi P, Kazemi E, Imielinski M, Elemento O, Hajirasouliha I. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. EBioMedicine 2018 Jan;27:317-328 [FREE Full text] [doi: 10.1016/j.ebiom.2017.12.026] [Medline: 29292031]
79. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 01;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]
80. Simidjievski N, Bodnar C, Tariq I, Scherer P, Andres Terre H, Shams Z, et al. Variational autoencoders for cancer data integration: design principles and computational practice. Front Genet 2019;10:1205 [FREE Full text] [doi: 10.3389/fgene.2019.01205] [Medline: 31921281]
81. Suresh H, Hunt N, Johnson A, Celi L, Szolovits P, Ghassemi M. Clinical intervention prediction and understanding with deep neural networks. In: Cornell University. 2017 Presented at: Machine Learning for Healthcare; 2017; Boston, Massachusetts, USA p. 322-337 URL: http://proceedings.mlr.press/v68/suresh17a.html
82. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. Bioinformatics 2019 Jul 15;35(14):i446-i454 [FREE Full text] [doi: 10.1093/bioinformatics/btz342] [Medline: 31510656]
83. Wang X, Sontag D, Wang F. Unsupervised learning of disease progression models. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. 2014 Presented at: KDD '14; 2014; New York, USA p. 85-94. [doi: 10.1145/2623330.2623754]
84. Gong J, Naumann T, Szolovits P, Guttag J. Predicting clinical outcomes across changing electronic health record systems. 2017 Presented at: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2017; Halifax, Nova Scotia, Canada p. 1497-1505 URL: http://library.usc.edu.ph/ACM/KKD%202017/pdfs/p1497.pdf [doi: 10.1145/3097983.3098064]
85. Defossez G, Rollet A, Dameron O, Ingrand P. Temporal representation of care trajectories of cancer patients using data from a regional information system: an application in breast cancer. BMC Med Inform Decis Mak 2014 Apr 02;14:24 [FREE Full text] [doi: 10.1186/1472-6947-14-24] [Medline: 24690482]
86. Mathew JP, Taylor BS, Bader GD, Pyarajan S, Antoniotti M, Chinnaiyan AM, et al. From bytes to bedside: data integration and computational biology for translational cancer research. PLoS Comput Biol 2007 Feb 23;3(2):e12 [FREE Full text] [doi: 10.1371/journal.pcbi.0030012] [Medline: 17319736]
87. Ramsay JO. Functional Data Analysis. New York: Springer-Verlag; 2005:4.

88. Alaa AM, van der Schaar M. Attentive state-space modeling of disease progression. In Advances in Neural Information Processing Systems. 2019. URL: https://papers.nips.cc/paper/2019/hash/1d0932d7f57ce74d9d9931a2c6db8a06-Abstract.html [accessed 2020-11-24]
89. Caravagna G, Graudenzi A, Ramazzotti D, Sanz-Pamplona R, De Sano L, Mauri G, et al. Algorithmic methods to infer the evolutionary trajectories in cancer progression. Proc Natl Acad Sci U S A 2016 Jul 12;113(28):E4025-E4034 [FREE Full text] [doi: 10.1073/pnas.1520213113] [Medline: 27357673]
90. Chung J, Kastner K, Dinh L, Goel K, Courville A, Bengio Y. A recurrent latent variable model for sequential data. In Advances in neural information processing systems. 2015. URL: https://papers.nips.cc/paper/2015/hash/b618c3210e934362ac261db280128c22-Abstract.html [accessed 2020-11-24]
91. Pereira S, Fontes F, Sonin T, Dias T, Fragoso M, Castro-Lopes JM, et al. Chemotherapy-induced peripheral neuropathy after neoadjuvant or adjuvant treatment of breast cancer: a prospective cohort study. Support Care Cancer 2016 Apr;24(4):1571-1581. [doi: 10.1007/s00520-015-2935-y] [Medline: 26384827]
92. Marshall TF, Zipp GP, Battaglia F, Moss R, Bryan S. Chemotherapy-induced-peripheral neuropathy, gait and fall risk in older adults following cancer treatment. J Cancer Res Pract 2017 Dec;4(4):134-138. [doi: 10.1016/j.jcrpr.2017.03.005]
93. Duregon F, Vendramin B, Bullo V, Gobbo S, Cugusi L, Di Blasio A, et al. Effects of exercise on cancer patients suffering chemotherapy-induced peripheral neuropathy undergoing treatment: A systematic review. Crit Rev Oncol Hematol 2018 Jan;121:90-100. [doi: 10.1016/j.critrevonc.2017.11.002] [Medline: 29198853]
94. Ospina P, McComb A, Wiart L, Eisenstat D, McNeely M. Physical therapy interventions, other than general physical exercise interventions, in children and adolescents before, during and following treatment for cancer. Cochrane Database Syst Rev 2018:2924. [doi: 10.1002/14651858.cd012924]
95. Axenie C, Becher A, Kurz D, Grauschopf T. Meta-Learning for Avatar Kinematics Reconstruction in Virtual Reality Rehabilitation. 2019 Presented at: IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE); October 28-30, 2019; Athens, Greece p. 617-624. [doi: 10.1109/bibe.2019.00117]
96. Streckmann F, Balke M, Lehmann HC, Rustler V, Koliamitra C, Elter T, et al. The preventive effect of sensorimotor- and vibration exercises on the onset of Oxaliplatin- or vinca-alkaloid induced peripheral neuropathies-STOP. BMC Cancer 2018 Jan 10;18(1):62 [FREE Full text] [doi: 10.1186/s12885-017-3866-4] [Medline: 29316888]
97. McCrary JM, Goldstein D, Sandler CX, Barry BK, Marthick M, Timmins HC, et al. Exercise-based rehabilitation for cancer survivors with chemotherapy-induced peripheral neuropathy. Support Care Cancer 2019 Feb 12;27(10):3849-3857. [doi: 10.1007/s00520-019-04680-w] [Medline: 30756229]
98. Zahiri M, Chen KM, Zhou H, Nguyen H, Workeneh BT, Yellapragada SV, et al. Using wearables to screen motor performance deterioration because of cancer and chemotherapy-induced peripheral neuropathy (CIPN) in adults - Toward an early diagnosis of CIPN. J Geriatr Oncol 2019 Nov;10(6):960-967 [FREE Full text] [doi: 10.1016/j.jgo.2019.01.010] [Medline: 30665876]
99. Schwenk M, Grewal GS, Holloway D, Muchna A, Garland L, Najafi B. Interactive sensor-based balance training in older cancer patients with chemotherapy-induced peripheral neuropathy: a randomized controlled trial. Gerontology 2016;62(5):553-563 [FREE Full text] [doi: 10.1159/000442253] [Medline: 26678611]
100. Grewal G, Sayeed R, Schwenk M, Bharara M, Menzies R, Talal TK, et al. Balance rehabilitation: promoting the role of virtual reality in patients with diabetic peripheral neuropathy. J Am Podiatr Med Assoc 2013;103(6):498-507. [doi: 10.7547/1030498] [Medline: 24297986]
101. Axenie C, Kurz D. Role of kinematics assessment and multimodal sensorimotor training for motion deficits in breast cancer chemotherapy-induced polyneuropathy: a perspective on virtual reality avatars. Front Oncol 2020;10:1419 [FREE Full text] [doi: 10.3389/fonc.2020.01419] [Medline: 32903366]
102. Bruzzone L, Chi M, Marconcini M. A novel transductive SVM for semisupervised classification of remote-sensing images. IEEE Trans Geosci Remote Sensing 2006 Nov;44(11):3363-3373. [doi: 10.1109/tgrs.2006.877950]
103. Kouroubali A, Koumakis L, Kondylakis H, Katehakis D. An integrated approach towards developing quality mobile health apps for cancer. In Mobile Health Applications for Quality Healthcare Delivery. 2019. URL: http://users.ics.forth.gr/~katehaki/publications/igibook2019.pdf [accessed 2020-11-24]
104. Kondylakis H, Koumakis L, Hänold S, Nwankwo I, Forgó N, Marias K, et al. Donor's support tool: Enabling informed secondary use of patient's biomaterial and personal data. Int J Med Inform 2017 Jan;97:282-292. [doi: 10.1016/j.ijmedinf.2016.10.019] [Medline: 27919386]
105. Belenkaya R, Campbell W, Chen R. Data standardization in cancer: challenges and opportunities. IMIA 2018.
106. Health Level Seven International (2020). URL: http://www.hl7.org/ [accessed 2020-11-21]
107. Personal Connected Health Alliance (2020). URL: http://www.pchalliance.org/ [accessed 2020-11-21]
108. SMART Health IT. URL: https://smarthealthit.org/
109. FHIR Release 4. HL7. URL: http://www.hl7.org/fhir/ [accessed 2020-11-21]
110. Mandel J, Kreda D, Mandl K, Kohane I, Ramoni R. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]

111. Warner J, Rioth M, Mandl K, Mandel JC, Kreda DA, Kohane IS, et al. SMART precision cancer medicine: a FHIR-based app to provide genomic information at the point of care. J Am Med Inform Assoc 2016 Jul;23(4):701-710 [FREE Full text] [doi: 10.1093/jamia/ocw015] [Medline: 27018265]
112. Triantafyllidis A, Kondylakis H, Votis K, Tzovaras D, Maglaveras N, Rahimi K. Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: A review of systematic reviews. Int J Med Inform 2019 Dec;132:103984. [doi: 10.1016/j.ijmedinf.2019.103984] [Medline: 31605884]
113. Kussaibi H, Macary F, Kennedy M, Booker D, Brodsky V, Schrader T, et al. HL7 CDA implementation guide for structured anatomic pathology reports methodology and tools. Stud Health Technol Inform 2010;160(Pt 1):289-293. [Medline: 20841695]
114. National Center for Chronic Disease PreventionHealth Promotion, Division of Cancer PreventionControl (2014). Implementation Guide for Ambulatory Healthcare Provider Reporting to Central Cancer Registries. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/phin/resources/guides/documents/implementation_guide_for_ambulatory_healthcare_provider_reporting_to_central_cancer_registries_march_2014.pdf [accessed 2020-11-21]
115. HL7 CDA ® Release 2 Implementation Guide: Reporting to Public Health Cancer Registries from Ambulatory Healthcare Providers, Release 1, DSTU Release 1. HL7 International (2015). 2014. URL: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=398 [accessed 2020-11-26]
116. McCabe RM. National cancer database: the past, present, and future of the cancer registry and its efforts to improve the quality of cancer care. Semin Radiat Oncol 2019 Oct;29(4):323-325. [doi: 10.1016/j.semradonc.2019.05.005] [Medline: 31472733]
117. Gong M, Wang Z, Liu Y. Toward early diagnosis decision support for breast cancer: ontology-based semantic interoperability. J Clin Oncol 2019;35:15. [doi: 10.1200/JCO.2019.37.15_SUPPL.E18072]
118. Chambers DA, Amir E, Saleh RR, Rodin D, Keating NL, Osterman TJ, et al. The impact of big data research on practice, policy, and cancer care. Am Soc Clin Oncol Educ Book 2019 Jan;39(39):e167-e175 [FREE Full text] [doi: 10.1200/EDBK_238057] [Medline: 31099675]
119. Sicotte C, Clavel S, Fortin M. A cancer care electronic medical record highly integrated into clinicians' workflow: users' attitudes pre-post implementation. Eur J Cancer Care (Engl) 2017 Nov 25;26(6):e12548. [doi: 10.1111/ecc.12548] [Medline: 27454341]
120. Improving Cancer-Related Outcomes with Connected Health: A Report to the President of the United States from the President's Cancer Panel. Bethesda (MD): President's Cancer Panel; 2016. URL: https://prescancerpanel.cancer.gov/report/connectedhealth/ [accessed 2020-11-24]
121. Study on Big Data in Public Health, Telemedicine and Healthcare. European Commision. 2016. URL: https://ec.europa.eu/digital-single-market/en/news/study-big-data-public-health-telemedicine-and-healthcare [accessed 2020-11-24]
122. Say RE, Thomson R. The importance of patient preferences in treatment decisions--challenges for doctors. Br Med J 2003 Sep 06;327(7414):542-545 [FREE Full text] [doi: 10.1136/bmj.327.7414.542] [Medline: 12958116]
123. Hlatky M. Patient preferences and clinical guidelines. J Am Med Assoc 1995 Apr 19;273(15):1219-1220. [Medline: 7707631]
124. Chong CA, Chen I, Naglie G, Krahn MD. How well do guidelines incorporate evidence on patient preferences? J Gen Intern Med 2009 Aug 23;24(8):977-982 [FREE Full text] [doi: 10.1007/s11606-009-0987-8] [Medline: 19387746]
125. Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. J Am Med Assoc 2008 Jul 23;300(4):436-438. [doi: 10.1001/jama.300.4.436] [Medline: 18647988]
126. Bastemeijer CM, Voogt L, van Ewijk JP, Hazelzet JA. What do patient values and preferences mean? A taxonomy based on a systematic review of qualitative papers. Patient Educ Couns 2017 May;100(5):871-881. [doi: 10.1016/j.pec.2016.12.019] [Medline: 28043713]
127. Luckmann R. Evidence-Based Medicine: How to Practice and Teach EBM, 2nd Edition: By David L. Sackett, Sharon E. Straus, W. Scott Richardson, William Rosenberg, and R. Brian Haynes, Churchill Livingstone, 2000. J Intensive Care Med 2016 Dec 07;16(3):155-156. [doi: 10.1177/088506660101600307]
128. Noble S, Matzdorff A, Maraveyas A, Holm MV, Pisa G. Assessing patients' anticoagulation preferences for the treatment of cancer-associated thrombosis using conjoint methodology. Haematologica 2015 Nov 20;100(11):1486-1492 [FREE Full text] [doi: 10.3324/haematol.2015.127126] [Medline: 26294737]
129. Froberg DG, Kane RL. Methodology for measuring health-state preferences--III: population and context effects. J Clin Epidemiol 1989;42(6):585-592. [doi: 10.1016/0895-4356(89)90155-8] [Medline: 2661731]
130. de Clercq PA, Blom JA, Korsten HM, Hasman A. Approaches for creating computer-interpretable guidelines that facilitate decision support. Artif Intell Med 2004 May;31(1):1-27. [doi: 10.1016/j.artmed.2004.02.003] [Medline: 15182844]
131. A new system for grading recommendations in evidence based guidelines. Br Med J. URL: https://www.bmj.com/content/323/7308/334.1 [accessed 2020-11-21]
132. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, GRADE Working Group. Grading quality of evidence and strength of recommendations. Br Med J 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]

133. NCCN Clinical Practice Guidelines in Oncology. URL: https://www.nccn.org/professionals/physician_gls/default.aspx [accessed 2020-11-21]
134. ICHOM: International Consortium for Health Outcomes Measurement. URL: http://www.ichom.org/ [accessed 2020-11-21]
135. Sackett DL, Rosenberg WM, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. Br Med J 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
136. Thomas L. Clinical practice guidelines. Evid Based Nurs 1999 Apr 01;2(2):38-39. [doi: 10.1136/ebn.2.2.38]
137. Tunis SR, Hayward RS, Wilson MC, Rubin HR, Bass EB, Johnston M, et al. Internists' attitudes about clinical practice guidelines. Ann Intern Med 1994 Jun 01;120(11):956-963. [doi: 10.7326/0003-4819-120-11-199406010-00008] [Medline: 8172440]
138. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PC, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. J Am Med Assoc 1999 Oct 20;282(15):1458-1465. [doi: 10.1001/jama.282.15.1458] [Medline: 10535437]
139. Liem EB, Obeid JS, Shareck EP, Sato L, Greenes RA. Representation of clinical practice guidelines through an interactive World-Wide-Web interface. Proc Annu Symp Comput Appl Med Care 1995:223-227 [FREE Full text] [Medline: 8563273]
140. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. J Am Med Assoc 2005 Mar 09;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
141. Wang D, Peleg M, Tu SW, Boxwala AA, Greenes RA, Patel VL, et al. Representation primitives, process models and patient data in computer-interpretable clinical practice guidelines:. Int J Med Inform 2002 Dec;68(1-3):59-70. [doi: 10.1016/s1386-5056(02)00065-5]
142. Liberman L, Menell JH. Breast imaging reporting and data system (BI-RADS). Radiol Clin North Am 2002 May;40(3):409-30, v. [doi: 10.1016/s0033-8389(01)00017-3] [Medline: 12117184]
143. Latoszek-Berendsen A, Tange H, van den Herik HJ, Hasman A. From clinical practice guidelines to computer-interpretable guidelines. A literature overview. Methods Inf Med 2010;49(6):550-570. [doi: 10.3414/ME10-01-0056] [Medline: 21085744]
144. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. Br Med J 1999 Feb 20;318(7182):527-530 [FREE Full text] [doi: 10.1136/bmj.318.7182.527] [Medline: 10024268]
145. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet 1993 Nov;342(8883):1317-1322. [doi: 10.1016/0140-6736(93)92244-n]
146. Beatty PA. Coping with abundance: the burden of progress in medical oncology. Oncologist 2012 Feb 06;17(2):294-295 [FREE Full text] [doi: 10.1634/theoncologist.2011-0230] [Medline: 22367600]
147. Muro N, Larburu N, Torres J, Kerexeta J, Artola G, Arrue M, et al. Architecture for a Multimodal and Domain-Independent Clinical Decision Support System Software Development Kit. In: Annu Int Conf IEEE Eng Med Biol Soc. 2019 Jul Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany, Germany p. 1399-1404 URL: https://ieeexplore.ieee.org/abstract/document/8856459 [doi: 10.1109/EMBC.2019.8856459]
148. Burki T. Vaccine misinformation and social media. Lancet Digit Health 2019 Oct;1(6):e258-e2e9. [doi: 10.1016/s2589-7500(19)30136-0]
149. Hurwitz B. Legal and political considerations of clinical practice guidelines. Br Med J 1999 Mar 06;318(7184):661-664 [FREE Full text] [doi: 10.1136/bmj.318.7184.661] [Medline: 10066215]
150. Jorritsma W, Cnossen F, van Ooijen P. Improving the radiologist-CAD interaction: designing for appropriate trust. Clin Radiol 2015 Feb;70(2):115-122. [doi: 10.1016/j.crad.2014.09.017] [Medline: 25459198]
151. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 2007;31(4-5):198-211 [FREE Full text] [doi: 10.1016/j.compmedimag.2007.02.002] [Medline: 17349778]
152. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? Radiology 2009 Oct;253(1):17-22. [doi: 10.1148/radiol.2531090689] [Medline: 19789251]
153. Rampun A, Wang H, Scotney B, Morrow P, Zwiggelaar R. Confidence Analysis for Breast Mass Image Classification. 2018 Presented at: 25th IEEE International Conference on Image Processing (ICIP); 2018; Athens p. 7-10. [doi: 10.1109/ICIP.2018.8451782]
154. Rampun A, Wang H, Scotney B, Morrow P, Zwiggelaar R. Classification of Mammographic Microcalcification Clusters with Machine Learning Confidence Levels. 2018 Jul Presented at: 14th International Workshop on Breast Imaging (IWBI); 2018; Atlanta, USA p. 8-11. [doi: 10.1117/12.2318058]
155. Lee RS, Gimenez F, Hoogi A, Rubin D. Curated Breast Imaging Subset of DDSM. The Cancer Imaging Archive. 2016. URL: https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM [accessed 2020-11-26]
156. Cahan A, Cimino JJ. A learning health care system using computer-aided diagnosis. J Med Internet Res 2017 Mar 08;19(3):e54 [FREE Full text] [doi: 10.2196/jmir.6663] [Medline: 28274905]
157. Artificial Intelligence and Machine Learning in Software as a Medical Device. US Food and Drug Administrations. URL: http://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device [accessed 2020-11-21]

158. Crico C, Renzi C, Graf N, Buyx A, Kondylakis H, Koumakis L, et al. mHealth and telemedicine apps: in search of a common regulation. Ecancermedicalscience 2018 Jul 11;12:853 [FREE Full text] [doi: 10.3332/ecancer.2018.853] [Medline: 30079115]
159. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a medical Device (SaMD). US Food and Drug Administration. URL: https://www.fda.gov/media/122535/download [accessed 2020-11-21]
160. Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. NPJ Digit Med 2019;2:69 [FREE Full text] [doi: 10.1038/s41746-019-0148-3] [Medline: 31372505]
161. Artificial Intelligence in healthcare: is Europe ready? Federation of European Academies of Medicine: FEAM. URL: https://www.feam.eu/wp-content/uploads/AI-Summary-report-15-Apr-2019-FV.pdf [accessed 2020-11-21]
162. Stratigi M, Kondylakis H, Stefanidis K. Fairgrecs: Fair group recommendations by exploiting personal health information. 2018 Presented at: International Conference on Database and Expert Systems Applications; August 9, 2018; Regensburg, Germany. [doi: 10.1007/978-3-319-98812-2_11]
163. Stratigi M, Kondylakis H, Stefanidis K. Fairness in group recommendations in the health domain. 2017 Presented at: IEEE 33rd International Conference on Data Engineering (ICDE); 2017; San Diego, USA URL: https://core.ac.uk/download/pdf/250169717.pdf [doi: 10.1109/ICDE.2017.217]

Abbreviations

ACC: accuracy
AI: artificial intelligence
AUC: area under the curve
CDSS: clinical decision support system
CIG: computer-interpretable guideline
CPG: clinical practice guideline
DSS: decision support system
EHR: electronic health record
ePRO: electronic patient-reported outcome
EU: European Union
FDA: Food and Drug Administration
H2020: Horizon 2020
ICT: information and communication technology
IEEE: Institute of Electrical and Electronics Engineers
IT: information technology
LSTM: long short-term memory
MD5: message-digest algorithm 5
mHealth: mobile health
ML: machine learning
NCCN: National Comprehensive Cancer Network
PROM: patient-reported outcome
SaMD: software as a medical device

Edited by G Eysenbach; submitted 01.07.20; peer-reviewed by T Exarchos, Y Xu; comments to author 08.08.20; revised version received 02.10.20; accepted 26.10.20; published 15.12.20.

Please cite as:
Kondylakis H, Axenie C, (Kiran) Bastola D, Katehakis DG, Kouroubali A, Kurz D, Larburu N, Macía I, Maguire R, Maramis C, Marias K, Morrow P, Muro N, Núñez-Benjumea FJ, Rampun A, Rivera-Romero O, Scotney B, Signorelli G, Wang H, Tsiknakis M, Zwiggelaar R
Status and Recommendations of Technological and Data-Driven Innovations in Cancer Care: Focus Group Study
J Med Internet Res 2020;22(12):e22034
URL: https://www.jmir.org/2020/12/e22034
doi:10.2196/22034
PMID:33320099

Original Paper

Social Network Analysis of Intensive Care Unit Health Care Professionals Measured by Wearable Sociometric Badges: Longitudinal Observational Study

Eiji Kawamoto[1,2], MD, PhD; Asami Ito-Masui[1,2], MD; Ryo Esumi[1,2], MD; Mami Ito[3], RS, MSc; Noriko Mizutani[3], RS; Tomoyo Hayashi[3], RS; Hiroshi Imai[1], MD, PhD; Motomu Shimaoka[2], MD, PhD

[1]Department of Emergency and Disaster Medicine, Graduate School of Medicine, Mie University, Tsu-City, Japan
[2]Department of Molecular Pathobiology and Cell Adhesion Biology, Graduate School of Medicine, Mie University, Tsu-City, Japan
[3]Emergency and Critical Care Center, Mie University Hospital, Tsu-City, Japan

Corresponding Author:
Motomu Shimaoka, MD, PhD
Department of Molecular Pathobiology and Cell Adhesion Biology
Graduate School of Medicine
Mie University
2-174 Edobashi Tsu-City
Edobashi
Tsu-City, 514-8507
Japan
Phone: 81 0592315031
Email: motomushimaoka@gmail.com

Abstract

Background: Use of wearable sensor technology for studying human teamwork behavior is expected to generate a better understanding of the interprofessional interactions between health care professionals.

Objective: We used wearable sociometric sensor badges to study how intensive care unit (ICU) health care professionals interact and are socially connected.

Methods: We studied the face-to-face interaction data of 76 healthcare professionals in the ICU at Mie University Hospital collected over 4 weeks via wearable sensors.

Results: We detail the spatiotemporal distributions of staff members' inter- and intraprofessional active face-to-face interactions, thereby generating a comprehensive visualization of who met whom, when, where, and for how long in the ICU. Social network analysis of these active interactions, concomitant with centrality measurements, revealed that nurses constitute the core members of the network, while doctors remain in the periphery.

Conclusions: Our social network analysis using the comprehensive ICU interaction data obtained by wearable sensors has revealed the leading roles played by nurses within the professional communication network.

(J Med Internet Res 2020;22(12):e23184) doi:10.2196/23184

KEYWORDS

wearable; interprofessional communication; clinician interaction; social network analysis

Introduction

Knowing how people interact in the workplace is an important basis for understanding how effective collaborative behavior develops. Social network analysis can be a potentially powerful tool for systematically assessing the nature of human interactions in the workplace, thereby elucidating the overall structure of organizational behavior. This can be depicted as a sociogram, which visualizes how interactions take place, how people are connected, how relationships are formed, and how information is transferred [1]. Social network analysis has been used to study human interactions in health care settings; however, its utility in promoting communication and collaboration between health care professionals remains to be demonstrated [2]. One of the factors potentially undermining social network analysis in health care and other relevant settings is the lack of an effective technology for objectively and comprehensively acquiring data [2]. Currently, datasets for social network analysis are usually acquired manually via questionnaires, observations, and manual retrieval of electronic medical and administrative records [3,4]. These manual data acquisition methods make it difficult to carry out objective and comprehensive/continuous measurements of social network connections among medical staff. The development of novel enabling technologies designed to address this problem has long been desired. Such technologies could improve the quality of social network analysis research by identifying the targets of interventions vis-à-vis collaborative behaviors.

Wearable sociometric sensor badges have emerged as a robust enabling technology to monitor such interactions objectively and comprehensively [5]. This study used wearable sensors, which have previously been used in both health care [6] and corporate [7] settings. They are embedded with 6 infrared data association transceivers on the front of the badge that detect interpersonal interactions. These wearable sensors also contain a built-in acceleration detector, thereby enabling the measurements of bodily motions associated with verbal communication. In addition, location information is obtained by using infrared beacons set at specific locations throughout the workplace. In this way, the wearable sensors are able to collect the data on who meets whom, when, where, and for how long, gathering objective and comprehensive datasets for social network analysis.

Effective interprofessional interactions are critical to the complex clinical environments involving multidisciplinary collaborations typical of intensive care units (ICUs) [8-10]. We recently performed a feasibility study [6] using wearable sociometric sensor badges in a health care setting to collect human behavioral interaction datasets. The design of our feasibility study involved the monitoring of face-to-face interactions among 76 health care professionals in the ICU at a university hospital in a comprehensive and objective manner. We demonstrated that the wearable sensor badges were able to objectively capture the temporal dynamics of the interactions between health care professionals at work. The intensities of the interactions peaked 3 times per day, exhibiting 8-hour interval periodic changes, which mirrored the reality of the ICU (ie, 3 shift changeovers spanning the day, evening, and night shifts). However, how the health care staff interacted interprofessionally in the ICU remains to be elucidated. Thus, building on our previous feasibility study, we carried out social network analysis to investigate the nature of the interprofessional interactions conducted in the ICU.

Methods

Participants

This study was approved by the Mie University Human Research Ethics Committee (approval no. 2978) as previously described in our feasibility study [6]. Briefly, a clinical investigation involving 76 ICU health care professionals was carried out in the ICU at Mie University Hospital. The ICU functions as a referral center for critical patients with such conditions as high-degree burns, post–cardiac arrest, septic shock, and other life-threatening traumas. Staff members working within the ICU (physicians, nurses, nursing assistants, pharmacists, medical technicians, desk clerks, and secretaries) were included in the study. Written consent was obtained from all participants. A few staff members who preferred not to participate wore mock badges that tuned off all sensor functions in order to maintain anonymity. A total of 76 medical professionals participated in the research project: 15 attending physicians, 39 nurses, 4 senior residents, 1 resident, 4 nursing assistants, 8 medical technicians, 2 receptionists, 1 pharmacist, and 2 secretaries. All participants wore the badges on the front of their clothing for 4 weeks during their working hours, including breaks.

Wearable Sensors and Data Acquisition

Wearable sociometric sensor badges (Business Microscope, Hitachi, Ltd, Tokyo) were used as previously described in our feasibility study of the ICU [6] and in another study of the corporate call center [7]. Briefly, the badges attached to the participants' front pockets enable automatic and comprehensive collection of the datasets required for the social network analysis of ICU health care professionals. These badges captured the wearer's physical movements via a 3-axis micro electro-mechanical acceleration sensor, which is used to detect individual activities. The badges also detect interpersonal interactions via 6 infrared data association transceivers on the front of the badge, as previously described [7]. Data on who met whom, when, and for how long can be collected. Location information is obtained by using infrared beacons set at specific locations within the ICU. Four infrared data association transceivers on the front of the beacon, facing at slightly different angles, create a detection range encompassing 60° horizontally and 30° vertically. The use of infrared beacons stationed in the ICU accords a 2-dimensional view of the location of the participants. With the use of these badges and the infrared beacons, the researchers were able to measure the duration and location of face-to-face communication among the study's participants. Two individuals were considered to have "actively communicated" with each other if there was a face-to-face event between them that exceeded a predefined threshold (ie, 2 Hz) of more than 1 minute. Active face-to-face interactions were previously classified as comprising gesture-aided conversation [6,7,11,12]. To locate the position of these human interactions within the ICU, a total of 249 infrared beacons were widely placed throughout the ICU as previously described [6] in 42 functional areas in which work-related interactions occurred, including 14 beds, a central nurses' station, conference room, consultation room, computed tomography scan control room, 4 examination-and-procedure rooms, family room, 3 laboratories, nurses' lounge, physicians' lounge, 2 physician stations, reception, satellite pharmacy, shower room, 4 utility rooms, and 5 storage rooms, but excluding corridors. As the infrared beacon beams, which are subject to interference by any concrete objects, travel up to 2 meters, multiple beacons were installed in each area to detect any interactions. Beacons between the nurses' station, bedsides, and other areas were far enough apart (ie, >4 m) to prevent any overlapping measurements.

Social Network Analysis

Using the datasets acquired by the wearable sensors, social network analysis was performed as previously described [5,7]. Social network connections are defined as face-to-face interactions between specific subjects (nodes) and are depicted as nodes connected with lines. The distances between nodes are inversely proportional to the total lengths of the face-to-face interactions between specific subjects. Centrality measurements were carried out as previously described [13,14].

Results

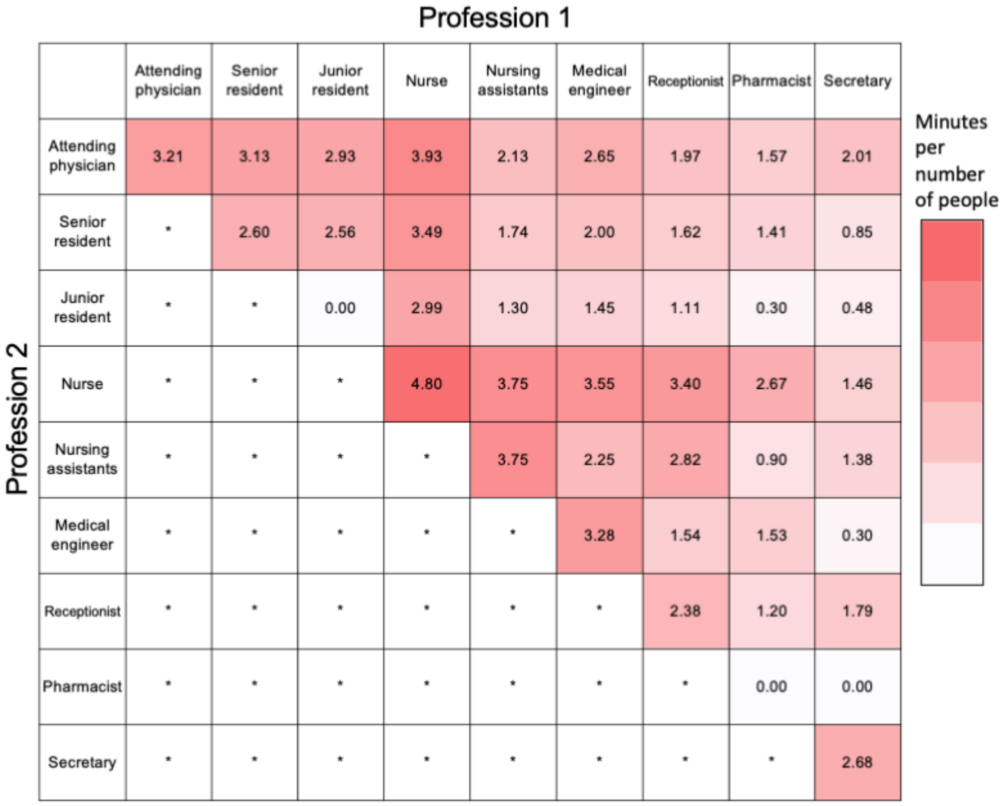
Spatiotemporal Distribution of the Active Face-to-Face Interactions

We analyzed a large data set previously collected in our earlier feasibility study that involved 76 ICU staff, each of whom worked for 160 hours on average during the 4-week period of data collection, totaling 729,600 minutes of active, person-to-person interaction. To determine exactly where the staff actively interacted with each other in the ICU ward each day, we mapped the daily-accumulated active interactions to the floor layout of the ICU during each day (from Days 1 to 30) (See Multimedia Appendix 1). The majority of the active interactions occurred in the patient area, in which several active interaction “hot spots” were identified at patient bedsides and the central nursing station. Outside of the patient area, hot spots were also observed in the nurses' lounge, the conference room, and the staff office (See Multimedia Appendix 2 and Multimedia Appendix 3).

Profession-Specific Active Communication

Next, we compared the total lengths of daily active communication per profession. The lengths of active daily interactions involving either nurses or nursing assistants were significantly greater than those of other professions (Figure 1). Nurses most often actively interacted with other nurses, and the length of their active interactions were greater than all of the other inter- and intraprofessional active interactions. Nurses also actively interacted well with many other professions (Figure 1). By contrast, nursing assistants actively interacted with each other and with nurses to a significant degree, but not with doctors. Attending physicians actively interacted mostly with nurses, followed by other attending physicians and residents. These data suggested the pivotal roles played by nurses and their interprofessional communication in the ICU.

Figure 1. Pair preferences for active face-to-face interactions among the ICU staff. A heat map representation showing the cumulative lengths of the active face-to-face interactions (min x person) between specific pairs of professions. The data are displayed via a logarithmic scale that reflects the length of the interactions multiplied by the number of people. *Not applicable.

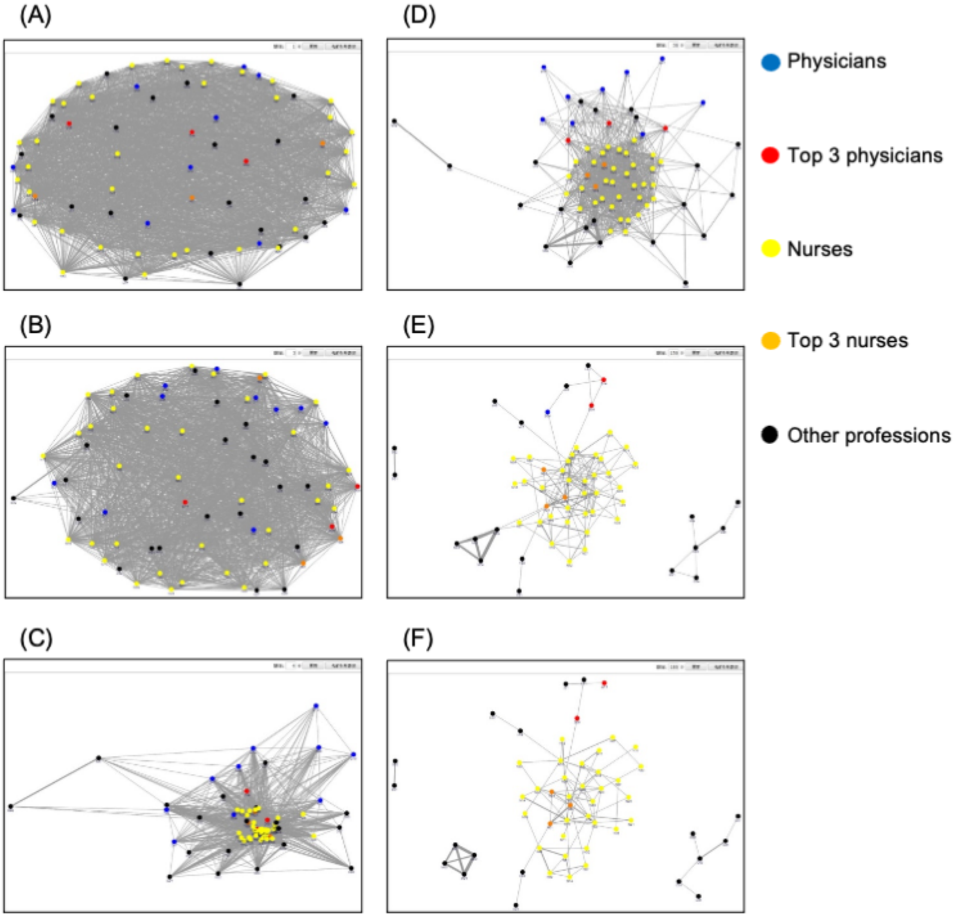


	Attending physician	Senior resident	Junior resident	Nurse	Nursing assistants	Medical engineer	Receptionist	Pharmacist	Secretary
Attending physician	3.21	3.13	2.93	3.93	2.13	2.65	1.97	1.57	2.01
Senior resident	*	2.60	2.56	3.49	1.74	2.00	1.62	1.41	0.85
Junior resident	*	*	0.00	2.99	1.30	1.45	1.11	0.30	0.48
Nurse	*	*	*	4.80	3.75	3.55	3.40	2.67	1.46
Nursing assistants	*	*	*	*	3.75	2.25	2.82	0.90	1.38
Medical engineer	*	*	*	*	*	3.28	1.54	1.53	0.30
Receptionist	*	*	*	*	*	*	2.38	1.20	1.79
Pharmacist	*	*	*	*	*	*	*	0.00	0.00
Secretary	*	*	*	*	*	*	*	*	2.68

Communication Network Analysis

To better understand the roles played by nurses in interprofessional communication in the ICU, we carried out a social network analysis. We focused our analysis on nurse-physician interactions, as nurses actively interacted not only with each other, but also with physicians (Figure 2). Moreover, nurses and physicians represent the two major groups in the ICU medical team [15]. The network connectivity between ICU staff was determined by the cumulative total lengths of active face-to-face interactions. By gradually changing the threshold from 1 to 180 min in order to determine the degree of connectivity (Figures 2A to 2F), different layers of the communication network in the ICU emerged. The communication network sociograms created using a threshold of 30 min (Figure 2D) demonstrated the core/periphery network structure [16], in which nurses were centrally positioned at the core, whereas physicians and other professions were at the periphery. The sociograms of the core group members contain many highly cohesive nodes and support the leadership roles played by nurses and their communication activity in the ICU, as previously shown by the social network analysis that mapped operating rooms [16].

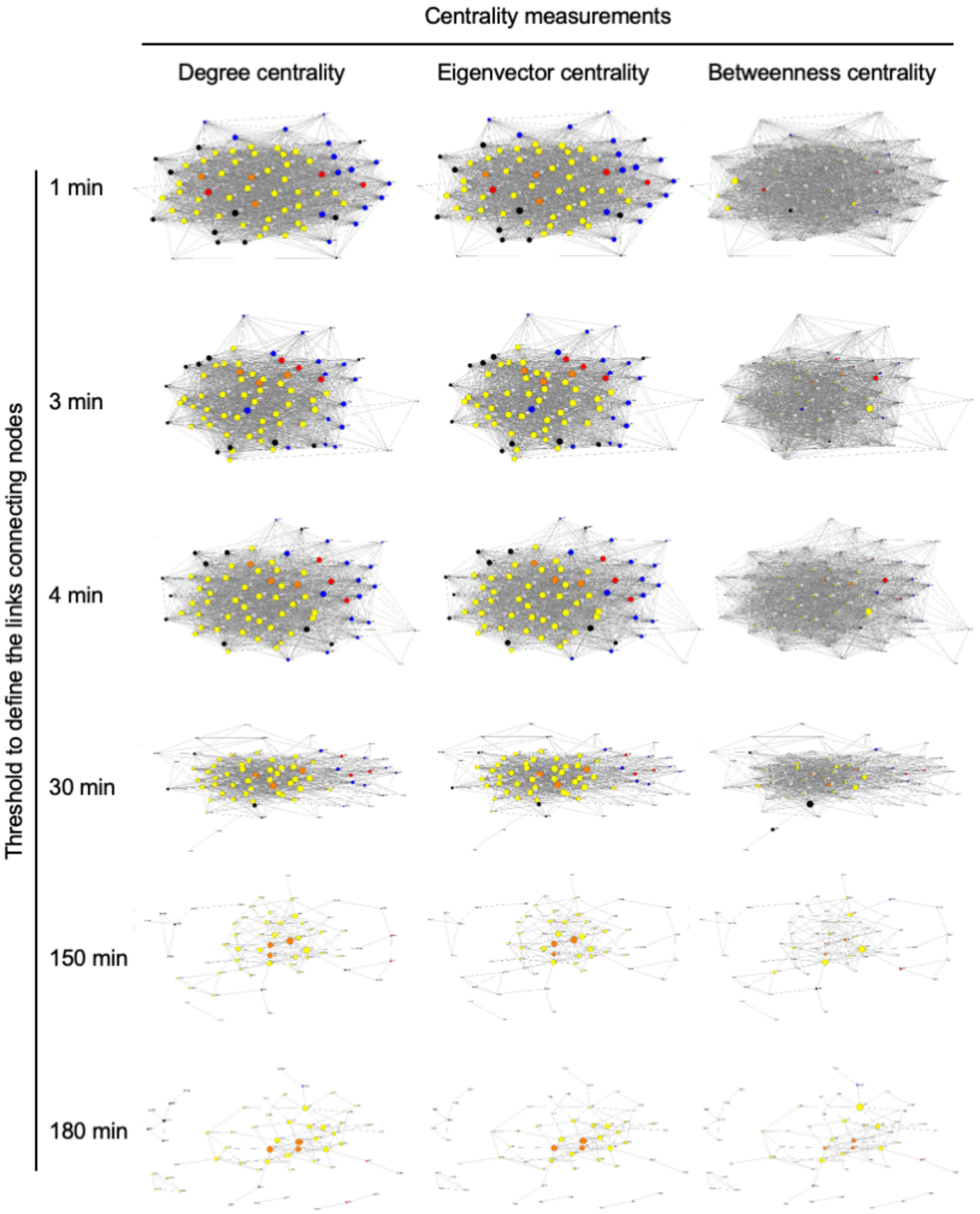
Figure 2. Social network analysis of active face-to-face interactions. Nodes represent individual ICU staff: blue and red, doctors; yellow and orange, nurses; black, other professions. The top 3 doctors and nurses who actively interacted the most are denoted by red and orange nodes, respectively. The following threshold values were used to define those links connecting the nodes: (A) 1, (B) 3, (C) 4, (D) 30, (E) 150, and (F) 180 minutes during the 4-week measurement period.



As the connectivity threshold rose to 150 and 180 min, focusing on highly communicative individuals in the ICU, a few non-nursing professions emerged at the periphery of the communication structure, whereas the core group remained a single cluster consisting of nurses (Figure 2E and Figure 2F). This suggests that high-volume communication in the ICU occurs predominantly as intraprofessional interactions among nurses.

To further substantiate our finding about the roles played by ICU nurses, we performed centrality measurements of the social network (Figure 3). We employed multiple centrality measurements such as the degree of centrality, eigenvector centrality, and betweenness centrality, in an attempt to identify the most prominent people in the network from different angles. By gradually changing the threshold from 1 to 180 min, different layers of the network in the ICU emerged, thereby demonstrating that nurses were centrally positioned at the core in all 3 centrality measurements (Figure 3), which confirmed the prominent roles of nurses in the communication network in the ICU.

Figure 3. Centrality measurements of the social network in the active face-to-face interactions between ICU healthcare professionals. Three different centrality measures (degree of centrality, eigenvector centrality, and betweenness centrality) were used. Nodes represent individual ICU staff: blue and red, doctors; yellow and orange, nurses; black, other professions. The top 3 doctors and nurses who actively interacted the most are denoted by red and orange nodes, respectively as shown in Figure 2. Threshold values used to define those links connecting the nodes are 1, 3, 4, 30, 150, and 180 minutes.



Discussion

By utilizing wearable sociometric sensor badges, the present study visualized the spatiotemporal distributions of the face-to-face interactions between health care professionals in the ICU. The social network analysis performed on the datasets collected by the wearable sensors also visualized the structures of the interprofessional interactions in the ICU, revealing the pivotal role played by nurses in the interaction network. The present study has extended our previous feasibility study of wearable sensors, which demonstrated not only that the interaction intensities changed periodically at 8-hour intervals, thereby reflecting the 3-shift changeover schedule of ICU workers (day, evening, and night shifts) [6], but also that the intensities of the interactions positively correlated with patient severity as measured by the APACHE II scores [6]. Similar sociometric wearable sensor technologies have been used to study the interactions between health care professionals, patients, and caregivers in a general pediatric hospital ward in Italy [17]; interactions between nurses at a surgical ICU at a university hospital in the US [18]; and interactions between doctors and nurses in the emergency department at university hospital in Italy [19]. In this way, the feasibility of the wearable sensors in monitoring human interactions in health care settings has been supported by multiple independent studies including ours [6].

Social network analysis has been used to study the relationships between individuals and the structures of teams in hospitals and other organizations [20]. Previous social network analysis studies performed in the field of multidisciplinary medical care have shown the important roles played by nurses. One study on operating room staffing identified a core/periphery network structure, in which nurses not only constituted the majority of the core members, but also led the multidisciplinary operating teams [16]. Another study on small ICU medical teams examined the formation of multidisciplinary communication networks, in which a doctor was at the center of one team network, while a nurse was at the center of another [21]. Yet another study on ICUs demonstrated that nurses represent an important group of influential individuals who create the central workplace culture [22]. The majority of previous social network analysis studies [16,20,21] used manual methods, such as surveys and observations, to collect connectivity data. The present study, which uses wearable sensors to continuously collect objective data, has substantiated the findings of previous social network analysis reports [16,20,21], namely that nurses play the leading role in such interprofessional communication networks. This conjuncture has been further corroborated by a series of centrality measurements of the network, including degree of centrality to describe the number of connections to a node; betweenness centrality to evaluate a node’s ability to connect others to one another; and eigenvector centrality to assess a node’s importance while simultaneously considering the importance of its neighboring nodes [13,14].

The nurse-to-physician ratio in ICUs might affect the patterns of social communication networks as more nurses could lead to their domination of communications. In the present study, the nurse-to-physician ratio in our ICU was 2.6, which appears to be higher than the estimated national average value of approximately 2 that we calculated from the publicly available data regarding ICUs in university hospitals in Japan. A previous study that examined 8 ICUs in university hospitals in France showed the nurse-to-physician ratio ranging from 1.8 to 5.6 [23]. How the patient-to-nurse and patient-to-physician ratios affect patient outcomes in ICUs has been well studied [23,24]. It is of continuing interest to study how the different nurse-to-physician ratios in ICUs facilitate such interprofessional collaborations, which can impact patient outcome.

Interprofessional communication and interactions are believed to play the pivotal role in facilitating the multidisciplinary collaborative work carried out in hospitals [25,26]. Effective collaborations between health care professionals may improve health care outcomes [25,26]. However, clinical evidence supporting the hypothesis that clinical and social interventions designed to promote interprofessional communication and collaboration do impact health care outcomes remains scarce [3]. One systematic review that examined 9 randomized clinical trials only found weak evidence supporting the claim that health care outcomes are improved by interventions to facilitate interprofessional collaborations [3]. It has been suggested that implementing robust analytic methods to examine human interactions is necessary if strong evidence in clinical trials is to be obtained [3]. Social network analysis might offer one such analytic method for studying the interprofessional interactions between health care professionals.

Social network analysis helps us better understand and objectively evaluate the structures of communication networks in ICUs, identifying targets of intervention that facilitate interprofessional communication. Some of our social network analysis results show that a few individuals play key roles in connecting the core members (nurses) to peripheral non-nursing professions, thereby acting as communication network hubs from which the flow of information is centered. By promoting such leading roles, highly effective interventions for facilitating wider interprofessional communication and collaboration can be developed and implemented. Alternatively, social network analysis could be used to objectively evaluate how interventions aimed at promoting interprofessional communication might actually improve these social communication networks.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science KAKENHI: grant number JP 18K08916 (HI), 19KK0224（EK), 18K08917 (EK), 19KK0196 (MS), 18H02622 (MS), 20K17859 (AI-M), 20K17858 (RE); and by the Japan Agency for Medical Research and Development (AMED): grant #19217687 (MS).

Conflicts of Interest

None declared.

Multimedia Appendix 1

The spatial distribution of active face-to-face interactions in the ICU. The daily sum of the lengths of the active interactions (persons X mins) was mapped to the floor layout of the ICU from days 1 to 30. Both inter- and intraprofessional interactions are included.

[MOV File , 871 KB - jmir_v22i12e23184_app1.mov]

Multimedia Appendix 2

The spatial distribution of active face-to-face interactions involving nurses in the ICU. The daily sum of the lengths of the active interactions (persons X mins) was mapped to the floor layout of the ICU from days 1 to 30. The nurse-to-nurse and nurse-to-other profession interactions are both included.
[MOV File , 871 KB - jmir_v22i12e23184_app2.mov]

Multimedia Appendix 3
The spatial distribution of active face-to-face interactions involving doctors in the ICU. The daily sum of the lengths of the active interactions (persons X mins) was mapped to the floor layout of the ICU from days 1 to 30. The doctor-to-doctor and doctor-to-other profession interactions are both included.
[MOV File , 870 KB - jmir_v22i12e23184_app3.mov]

References

1. Knoke D, Yang S. Social network analysis. Thousand Oaks: Sage Publications; 2020.
2. Chambers D, Wilson P, Thompson C, Harden M. Social network analysis in healthcare settings: a systematic scoping review. PLoS One 2012;7(8):e41911 [FREE Full text] [doi: 10.1371/journal.pone.0041911] [Medline: 22870261]
3. Reeves S, Pelone F, Harrison R, Goldman J, Zwarenstein M. Interprofessional collaboration to improve professional practice and healthcare outcomes. Cochrane Database Syst Rev 2017 Jun 22;6:CD000072 [FREE Full text] [doi: 10.1002/14651858.CD000072.pub3] [Medline: 28639262]
4. Walters SJ, Stern C, Robertson-Malt S. The measurement of collaboration within healthcare settings: a systematic review of measurement properties of instruments. JBI Database System Rev Implement Rep 2016 Apr;14(4):138-197. [doi: 10.11124/JBISRIR-2016-2159] [Medline: 27532315]
5. Ara K, Kanehira N, Olguín DO, Waber BN, Kim T, Mohan A, et al. Sensible Organizations: Changing Our Businesses and Work Styles through Sensor Data. Journal of Information Processing 2008;16:1-12. [doi: 10.2197/ipsjjip.16.1]
6. Ito-Masui A, Kawamoto E, Nagai Y, Takagi Y, Ito M, Mizutani N, et al. Feasibility of Measuring Face-to-Face Interactions among ICU Healthcare Professionals Using Wearable Sociometric Badges. Am J Respir Crit Care Med 2020 Jan 15;201(2):245-247. [doi: 10.1164/rccm.201904-0779LE] [Medline: 31545651]
7. Watanabe J, Ishibashi N, Yano K. Exploring relationship between face-to-face interaction and team performance using wearable sensor badges. PLoS One 2014;9(12):e114681 [FREE Full text] [doi: 10.1371/journal.pone.0114681] [Medline: 25501748]
8. Shoham DA, Mundt MP, Gamelli RL, McGaghie WC. The Social Network of a Burn Unit Team. J Burn Care Res 2015;36(5):551-557. [doi: 10.1097/BCR.0000000000000218] [Medline: 25501788]
9. Kelly Costa D, Liu H, Boltey EM, Yakusheva O. The Structure of Critical Care Nursing Teams and Patient Outcomes: A Network Analysis. Am J Respir Crit Care Med 2020 Feb 15;201(4):483-485. [doi: 10.1164/rccm.201903-0543LE] [Medline: 31626555]
10. Goldman J, Kitto S, Reeves S. Examining the implementation of collaborative competencies in a critical care setting: Key challenges for enacting competency-based education. J Interprof Care 2018 Jul;32(4):407-415. [doi: 10.1080/13561820.2017.1401987] [Medline: 29161170]
11. Watanabe JI, Fujita M, Yano K, Kanesaka H, Hasegawa T. Resting time activeness determines team performance in call centers. 2012 Presented at: international conference on social informatics IEEE; 2012; Alexandria, Virginia, USA p. 26-31. [doi: 10.1109/SocialInformatics.2012.40]
12. Ara K, Sato N, Takahashi K, Maeda H, Yano K. Integrating Wearable Sensor Technology into Project-management Process. Journal of Information Processing 2012;20(2):406-418 [FREE Full text] [doi: 10.2197/ipsjjip.20.406]
13. Landherr A, Friedl B, Heidemann J. A Critical Review of Centrality Measures in Social Networks. Bus Inf Syst Eng 2010 Oct 20;2(6):371-385. [doi: 10.1007/s12599-010-0127-3]
14. Shafiei SM, Azar A, Esmaielpour R. Mapping and Social Network Analysis of the Nurses of Razi Hospital. Iran Red Crescent Med J 2018 May 05;20(10):e58321. [doi: 10.5812/ircmj.58321]
15. Rose L. Interprofessional collaboration in the ICU: how to define? Nurs Crit Care 2011;16(1):5-10. [doi: 10.1111/j.1478-5153.2010.00398.x] [Medline: 21199549]
16. Anderson C, Talsma A. Characterizing the structure of operating room staffing using social network analysis. Nurs Res 2011;60(6):378-385. [doi: 10.1097/NNR.0b013e3182337d97] [Medline: 22048555]
17. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, Van DBW, et al. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. PLoS One 2011 Feb 28;6(2):e17144 [FREE Full text] [doi: 10.1371/journal.pone.0017144] [Medline: 21386902]
18. Rosen MA, Dietz AS, Lee N, Wang I, Markowitz J, Wyskiel RM, et al. Sensor-based measurement of critical care nursing workload: Unobtrusive measures of nursing activity complement traditional task and patient level indicators of workload to predict perceived exertion. PLoS One 2018;13(10):e0204819 [FREE Full text] [doi: 10.1371/journal.pone.0204819] [Medline: 30312326]

19. Stefanini A, Aloini D, Gloor P, Pochiero F. Patient satisfaction in emergency department: Unveiling complex interactions by wearable sensors. Journal of Business Research 2020:1-12 (forthcoming) [FREE Full text] [doi: 10.1016/j.jbusres.2019.12.038]
20. Sabot K, Wickremasinghe D, Blanchet K, Avan B, Schellenberg J. Use of social network analysis methods to study professional advice and performance among healthcare providers: a systematic review. Syst Rev 2017 Oct 23;6(1):208 [FREE Full text] [doi: 10.1186/s13643-017-0597-1] [Medline: 29058638]
21. Lurie SJ, Fogg TT, Dozier AM. Social network analysis as a method of assessing institutional culture: three case studies. Acad Med 2009 Aug;84(8):1029-1035 [FREE Full text] [doi: 10.1097/ACM.0b013e3181ad16d3] [Medline: 19638768]
22. Fong A, Clark L, Cheng T, Franklin E, Fernandez N, Ratwani R, et al. Identifying influential individuals on intensive care units: using cluster analysis to explore culture. J Nurs Manag 2017 Jul;25(5):384-391. [doi: 10.1111/jonm.12476] [Medline: 28568480]
23. Neuraz A, Guérin C, Payet C, Polazzi S, Aubrun F, Dailler F, et al. Patient Mortality Is Associated With Staff Resources and Workload in the ICU: A Multicenter Observational Study. Crit Care Med 2015 Aug;43(8):1587-1594. [doi: 10.1097/CCM.0000000000001015] [Medline: 25867907]
24. Penoyer DA. Nurse staffing and patient outcomes in critical care: a concise review. Crit Care Med 2010 Jul;38(7):1521-8; quiz 1529. [doi: 10.1097/CCM.0b013e3181e47888] [Medline: 20473146]
25. Tan T, Zhou H, Kelly M. Nurse-physician communication - An integrated review. J Clin Nurs 2017 Dec;26(23-24):3974-3989. [doi: 10.1111/jocn.13832] [Medline: 28370533]
26. Varpio L, Hall P, Lingard L, Schryer CF. Interprofessional communication and medical error: a reframing of research questions and approaches. Acad Med 2008 Oct;83(10 Suppl):S76-S81. [doi: 10.1097/ACM.0b013e318183e67b] [Medline: 18820507]

Abbreviations

ICU: intensive care unit

Edited by G Eysenbach; submitted 04.08.20; peer-reviewed by S Piva, E Alanazi, A Ramachandran; comments to author 25.09.20; revised version received 29.10.20; accepted 12.11.20; published 31.12.20.

Please cite as:
Kawamoto E, Ito-Masui A, Esumi R, Ito M, Mizutani N, Hayashi T, Imai H, Shimaoka M
Social Network Analysis of Intensive Care Unit Health Care Professionals Measured by Wearable Sociometric Badges: Longitudinal Observational Study
J Med Internet Res 2020;22(12):e23184
URL: http://www.jmir.org/2020/12/e23184/
doi:10.2196/23184
PMID:33258785

Review

Potential Benefits of Multimedia-Based Home Catheter Management Education in Patients With Peripherally Inserted Central Catheters: Systematic Review

Kija Malale[1*], BSc, MSc, PhD; Jili Fu[2,3*], MD, MSc; William Nelson[4], BSc, MSc; Helena Marco Gemuhay[5], BSc, MSc; Xiuni Gan[1], BSc, MSc, PhD; Zhechuan Mei[6], PhD

[1]Department of Nursing, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China
[2]Daping Hospital, Army Medical University, Chongqing, China
[3]Department of Endocrinology and Metabolism, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China
[4]School of Public Health and Social Sciences, Muhimbili University of Health and Allied Sciences, Dar es Salaam, United Republic of Tanzania
[5]School of Nursing, St John's University of Tanzania, Dodoma, United Republic of Tanzania
[6]Department of Gastroenterology, The Second Hospital of Chongqing Medical University, Chongqing, China
[*]these authors contributed equally

Corresponding Author:
Xiuni Gan, BSc, MSc, PhD
Department of Nursing
The Second Affiliated Hospital of Chongqing Medical University
76, Linjiang Road, Yuzhong District
Chongqing, 400010
China
Phone: 86 138 83309318
Email: ganxn@163.com

Abstract

Background: In recent years, there have been many suggestions to use multimedia as a strategy to fully meet the educational needs of patients with peripherally inserted central catheters. However, the potential benefits remain unreliable in the literature.

Objective: In this study, we identified the potential benefits of multimedia-based home catheter management education in patients with peripherally inserted central catheters and discussed the clinical implications.

Methods: We performed systematic searches of the PubMed, Cochrane Library, Embase Ovid, Medline, BioMed Central-cancer (BMC-cancer), ScienceDirect, and Google Scholar databases without date constraints until November 30, 2019. The methodological quality of the eligible studies was appraised using the Cochrane risk of bias tool. Narrative synthesis of the study findings was conducted.

Results: A total of 6 intervention studies met the inclusion criteria, including 3 randomized controlled trials and 3 case-control studies/quasi-experimental studies. The studies included a total of 355 subjects, including a total of 175 in the multimedia groups and 180 in the control groups. We identified 4 potential benefits to patients: (1) improved knowledge, (2) increased satisfaction, (3) reduced incidence of catheter-related complications, and (4) reduced number of cases of delayed care after complications.

Conclusions: The current systematic review highlights the potential benefits of multimedia-based home catheter management education for patients with peripherally inserted central catheters.

(J Med Internet Res 2020;22(12):e17899) doi:10.2196/17899

KEYWORDS

home catheter management; multimedia-based education; peripherally inserted central catheter

Introduction

Background

With the innovation of vascular access, the application of peripherally inserted central catheters is transforming care for patients treated with long-term infusion therapy in home care [1-3]. A peripherally inserted central catheter is a vascular access device that is inserted into the superficial or deep veins of the upper or lower extremities and advances the distal third of the superior vena cava or the proximal third of the inferior vena cava [1,2]. These catheters are often used in cancer patients, especially those requiring long-term infusion therapy [2]. Compared to the use of central venous catheters, infusion of vesicant/irritants and hypertonic solutions using peripherally inserted central catheters is safer, less costly, and more reliable [1]. During treatment, patients can live at home with a catheter for weeks to months [4,5].

Previous studies have shown that home catheter management education before catheterization plays an important role not only in preventing catheter-related complications, but also in improving catheter retention [6-9]. In order to maximize the clinical benefits of peripherally inserted central catheter services, home catheter management education is provided to each patient prior to catheterization. The information provided includes how to flush the catheter, when to change the dressing, when to clean the catheter, how to identify signs and symptoms of catheter-related complications, how to identify high-risk behaviors, and how to check if the catheter is inserted correctly [10,11]. However, the main challenge remaining is the lack of educational approaches that could effectively meet the educational needs of patients. The traditional didactic approach, commonly used in clinical settings around the world, is considered ineffective [1,3,12-16]. Patients complain that they get too little or too much information, which can be unhelpful, scary, technical, and hard to understand [12-14,16,17]. Patients whose educational needs are not met can become distressed, dissatisfied with the care provided, and have a diminished quality of life [18-22]. Sometimes they demand the catheter to be removed immediately after insertion [15]. In addition, other studies have found an increased incidence of catheter-related complications, including infection, catheter obstruction, thrombosis, and catheter displacement [23]. These hurdles stress the need for a new approach to effectively meet the educational needs of patients.

Technological innovation is not lagging in responding to the education challenges of patients, especially patients with peripherally inserted central catheters [3,6,9,10,23]. Patients benefit from the adaptability of technology and can use it to start, stop, and resume learning at their own pace. In recent years, some centers have experimented with technological innovations in the form of multimedia (eg, text, audio, images, animation, video, and voiceover interactive PowerPoint) to educate patients [9-11,24]. However, the potential benefits of multimedia-based education remain unclear. To our understanding, the existing literature is conflicting, which can confuse clinicians when choosing effective approaches to meet patients' educational needs [9-11,25]. Moreover, during our literature search, we found no systematic review or meta-analysis study that could synthesize the existing evidence.

Objective

The main objective of this review was to determine the potential benefits of multimedia-based home catheter management education in patients with peripherally inserted central catheters and to discuss the clinical implications.

Methods

Design

The review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines [26]. A detailed protocol for the study was formulated prior to data collection.

Search Strategy and Selection Criteria

The search strategy was conducted using MeSH key terms and their respective original words. We searched for the following: “Patient Education as Topic“ [Majr] OR “Patient Education” OR “Health Education” [Majr] OR “Health Education” AND “Catheterization, Peripheral” [Majr] OR “Peripherally Inserted Central Catheter” OR “PICC”. Electronic databases such as PubMed, Cochrane Library, Embase Ovid, Medline, BioMed Central-cancer (BMC-cancer), ScienceDirect, and Google Scholar were searched without date constraints until November 30, 2019. The review was limited to published articles written in English. Additional sources were identified by hand-searching reference lists of relevant study using Google search. Two independent researchers conducted the search in accordance with the set criteria.

Studies were included based on the following set of inclusion criteria: (1) an interventional study with two or more comparative groups; (2) involved subjects who had been prescribed or had already been installed with a peripherally inserted central catheter; (3) the educational approaches (interventions) employed were either didactic sessions (orally presented) as a standard (control) or were supplemented with multimedia (text, audio, image, animation, video, or voiceover interactive PowerPoint) for the test group; (4) the study reported one or more outcomes. Studies were excluded if (1) they included subjects aged <18 years (children); (2) they had unclear educational interventions; (3) subjects received interventions of a similar kind prior the study or different interventions during the follow-up period; or (4) they included critically ill patients (See Multimedia Appendix 1: PubMed search results summary).

Quality Appraisal

We used the Cochrane risk assessment tool [27] to assess the risk of bias in randomized controlled trials and the ROBIN-1 tool [28] to assess the risk of bias in nonrandomized studies. In randomized controlled trials, the risk of bias was judged to be low risk, high risk, or unclear risk in each of these 6 aspects: (1) random sequence generation (selection bias), (2) allocation concealment (selection bias), (3) blinding of participants and personnel (performance bias), (4) blinding of outcome assessment (detection bias), (5) risk of attrition and reporting bias, and (6) any other sources of bias. In the nonrandomized studies, the risk of bias in each study was judged to be low risk, moderate risk, serious risk, critical risk, or no information in each of these seven aspects: (1) bias due to confounding, (2) bias in selection of participants into the study, (3) bias in classification of interventions, (4) bias due to deviations from intended interventions, (5) bias due to missing data, (6) bias in measurement of outcomes, and (7) bias in selection of the reported results. Two independent evaluators assessed the methodological quality of each study. If their results were in conflict, a third evaluator was included. The strength of evidence per outcome was rated using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) software [29].

Data Abstraction and Synthesis

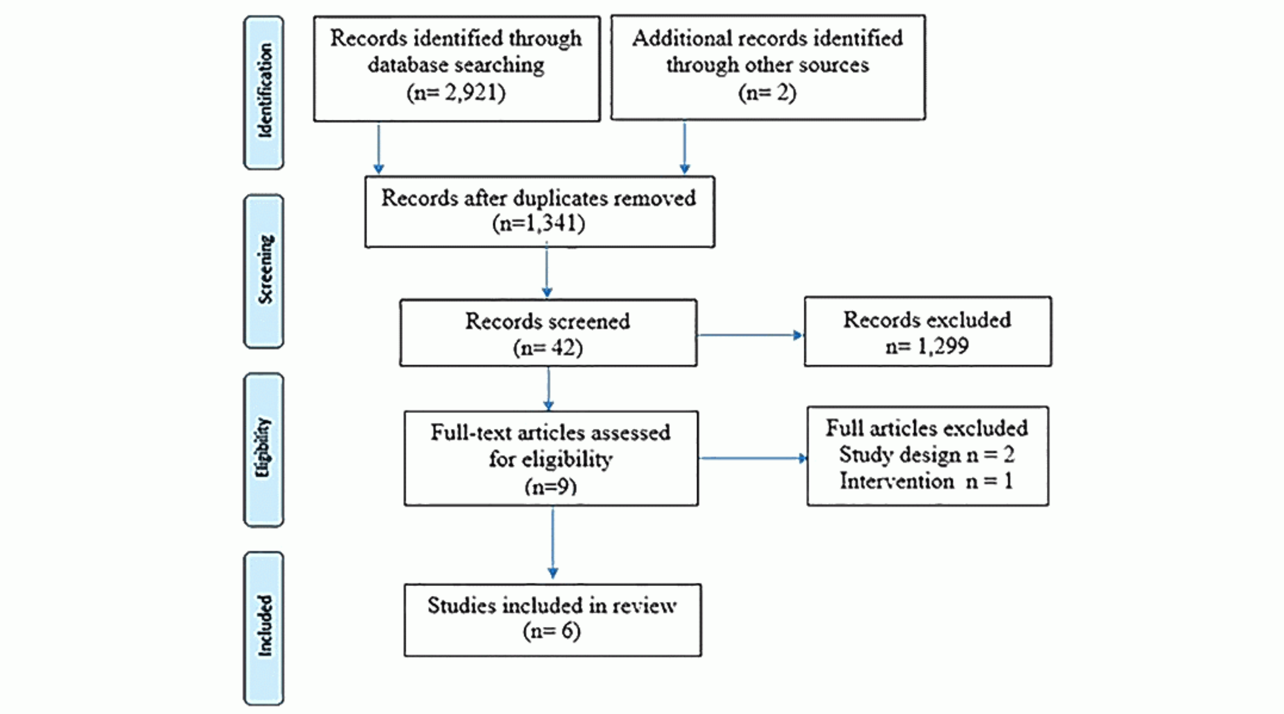
Data were extracted by 2 independent data collectors following a discussion of any conflicting results. Information pertaining to author name(s), year, country, study aim/outcomes, design, intervention type, sample size, and key findings in each study were extracted. Methodological quality of the included studies was uneven, which limited the meta-analysis. We chose to report the findings across the studies using a narrative synthesis format.

Results

Search Results

A search produced 2923 studies: PubMed (222), Cochrane Library (10), Embase Ovid (23), Medline (10), BMC-Cancer (1), ScienceDirect (281), and Google Scholar (2374). The other 2 studies come from the Google search engine. After removing duplicates and screening, 9 studies were considered potentially eligible [6-11,24,25,30]. After further evaluation, 6 studies met our inclusion criteria [8-11,24,25]. Two studies were excluded because of unclear educational interventions [6,7], and the other because of unrelated design [30].Three studies came from the United States [9,10,25], 1 from Canada [24], 1 from France [8], and 1 from Italy [11]. In addition, 3 studies were randomized controlled trials [9,11,24], and 3 were quasi-experimental/case-control studies [8,10,25]. Figure 1 shows the PRISMA flow chart for screening and selecting studies.

Figure 1. PRISMA flow chart for screening and selecting studies.



Study Characteristics

The main characteristics of the 6 eligible studies are shown in Table 1. These studies included a total of 355 subjects, including 175 (49%) in the multimedia-based education group. The number of subjects in each study ranged from 11 [10] to 130 [25]. Follow-up for each study ranged from 1 day to 10 months. Different studies use different multimedia formats. Of the 6 studies, 4 used multimedia in the form of video [10,11,24,25], 1 used voiceover interactive PowerPoint [9], and 1 used pictures [8]. Eligible studies reported various outcomes, including patient understanding [24], patient comprehension [8], patient knowledge [9,11,25], patient satisfaction [8,10,24,25], catheter-related complications [8-10], and seeking medical care after complications occurred [8,9]. For the purposes of the current review, the terms “patient understanding,” “patient knowledge,” and “patient understanding” are merged into “patient knowledge.” Of the 5 studies that reported on patient knowledge, 4 used questionnaires [9,11,24,25], and 1 used the repeat-back method [8]. In the studies, patient satisfaction was evaluated by either questionnaire [10,24], the Likert-scale [8,25], or observation [11]. Catheter-related complications and seeking medical care after complications occurred were assessed by counting the number of cases of complications and the number of calls to health care providers/the number cases of delayed medical care after complications occurred.

Table 1. Eligible study characteristics.

Authors, country	Outcome(s)	Design	Sample size, n (control sample size)	Interventions
Veyrier et al [8], France	Comprehension, satisfaction, occurrence of adverse events, and seeking medical care after adverse events	Prospective case-control	30 (17)	OP[a] plus cartoon and card game versus OP alone (control)
Fusco et al [11], Italy	Knowledge	Randomized controlled trial	40 (27)	OP plus video presentation versus OP alone (control)
Emery et al [9], USA	Knowledge, catheter-related complications, and seeking medical care	Randomized controlled trial	51 (27)	OP plus VOIPP[b] versus OP alone (control)
Bowers et al [24], Canada	Understanding and satisfaction	Randomized controlled trial	93 (44)	OP plus video presentation versus OP alone (control)
Petroulias [10], USA	Satisfaction and catheter-related complications	Quasi-experimental study	11 (national data)	OP plus video presentation versus OP alone (control)
Sowan et al [25], USA	Knowledge recall, retention, satisfaction with the consent process, and multimedia patient decision aids	Quasi-experimental study	130 (65)	OP plus multimedia patient decision aids versus OP alone (control)

[a] OP: oral presentation.

[b] VOIPP: voiceover interactive PowerPoint.

Methodological Quality of Eligible Studies

Randomized Controlled Trials

Of the 3 randomized controlled trials, the risk of random sequence generation (selection bias) was considered low in 2 studies [11,24] and unclear in 1 [9]. Only 1 study [24] involved allocation concealment (selection bias). The risk of bias due to the blinding of participants and personnel (performance bias) was unknown for all of the studies. Blinding of outcome assessment (detection bias) was considered high-risk in 2 studies [9,11] and unclear in 1 study [24]. All studies had a low risk of attrition and reporting bias. The other sources of bias were also considered low-risk in all studies. Table 2 shows risk of bias summary for randomized controlled trials.

Table 2. Risk of bias summary for randomized controlled trials.

Author	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Bowers et al [24]	✓	✓			✓	✓	
Emery et al [9]				×	✓	✓	
Fusco et al [11]	✓			×	✓	✓	

Nonrandomized Controlled Trials

Of the nonrandomized studies, 2 of the 3 [8,10] were identified as having serious confounding-bias, and 1 as having moderate confounding bias [25]. Two of the studies [8,25] had a moderate selection bias risk, and 1 [10] had a critical risk of bias. In 2 studies [8,10], the risk of bias in the intervention classification was considered moderate, while in the other study [25] it was considered low. The risk of bias due to missing data was considered low in 2 studies [10,25] and critical in the third study [8]. The risk of bias in outcome measures was considered low in 2 studies [10,25] and critical in the other. In addition, bias in selection of the reported results was considered low risk in all 3 studies. The overall risk of bias was considered critical in 2 studies [8,10] and moderate in the other study [25]. Table 3 shows risk of bias summary for nonrandomized controlled trials.

Table 3. Risk of bias summary for nonrandomized controlled trials.

Authors	Bias due to confounding	Bias in selection of participant	Bias in classification of intervention	Bias due to missing data	Bias in measurement of outcome	Bias in selection of the reported results	Overall results
Petroulias [10]	Serious	Critical	Moderate	low	Low	Low	Critical
Veyrier et al [8]	Serious	Moderate	Moderate	Critical	Moderate	Low	Critical
Sowan et al [25]	Moderate	Moderate	Low	Low	Low	Low	moderate

Synthesis

Overview

After a comprehensive narrative synthesis of the eligible key findings, 4 key aspects were identified as potential benefits of multimedia-based home catheter management education for patients with peripherally inserted central catheters. Key aspects identified include (1) knowledge improvement; (2) satisfaction with the information and services provided; (3) incidence of catheter-related complications; and (4) improved seeking of medical care after complications occur. The strength of evidence for each principal outcome was also measured (see Multimedia Appendix 2: Principal findings summary and evidence strength of each outcome).

Patient Knowledge Improvement

Five studies evaluated the knowledge of patients after intervention [8,9,11,24,25]. However, due to the different assessment tools and educational contents across studies, meta-analysis could not be conducted. Emery and colleagues [9] used a self-administered questionnaire containing 6 items. Items included (1) proper length of time to clean the end connector, (2) proper length of time to wash hands before initiating parenteral nutrition procedure, (3) preferred solution to clean work surfaces, (4) the first step to take when a crack or hole is found in a catheter, (5) steps that do not require notification of the Home Parenteral Nutrition service, and (6) important steps in preventing accidental removal of the catheter. Bowers and colleagues [24] used a self-administered questionnaire containing 5 true or false questions based on a particular procedure as well as the nature, risks, benefits, and procedures of catheter installation. Fusco [11] and colleagues used a self-administered questionnaire consisted of a series of questions, including closed, dichotomous, multiple choice, and open questions[11]. Veyrier and colleagues [8] used a repeat-back technique to assess how well patients understood what they had learned in previous discussions. Sowan and colleagues [25] used a structured questionnaire with 19 multiple choice and true or false questions covering the procedure indications, benefits, contraindications, insertion site, complications, and patient and health care team roles in the care and safety of a peripherally inserted central catheter. Despite differences in research methods, patient knowledge improved in all studies regardless of the educational intervention. Nevertheless, the improvement of knowledge in the multimedia groups was better than that in the control groups [8,11,24,25]. Table 4 summarizes the effects of multimedia-based education on patient knowledge improvement.

Table 4. Effect of multimedia-based education on patient knowledge improvement.

Author, country	Design	Interventions	Follow-up (days)	Assessment tool	Narrative summary of results
Veyrier et al [8], France	Prospective case-control	OP[a] plus cartoon and card game versus OP alone (control)	1-2	Repeat-back	The overall adverse events comprehension score of the test group was higher than that of the control group.
Fusco et al [11], Italy	Randomized controlled trial	OP plus video presentation versus OP alone (control)	1	Questionnaire	Compared with the OP group, the overall peripherally inserted central catheter management knowledge score of the test group improved more.
Emery et al [9], USA	Randomized controlled trial	OP plus VOIPP[b] versus OP alone (control)	1-10	Questionnaire	Catheter care knowledge scores increased in similar ways in each group immediately after the intervention and 7-10 days after the intervention.
Bowers et al [24], Canada	Randomized controlled trial	OP plus video presentation versus OP alone (control)	1	Questionnaire	The test group was significantly better than the control group in peripherally inserted central catheter insertion procedure comprehension score.
Sowan et al [25], USA	Quasi-experimental study	OP plus multimedia patient decision aids versus OP alone (control)	1-2	Questionnaire	Compared with the control group, the multimedia group scored 2 points higher in knowledge recall and retention with the consent process.

[a] OP: oral presentation.

[b] VOIPP: voiceover interactive PowerPoint.

Patient Satisfaction With the Information and Services Provided

Four studies assessed patient satisfaction with the information and services provided [8,10,24,25], but meta-analysis was not possible due to the use of different tools for evaluation. Veyrier and colleagues [8] used a 4-point Likert-type scale that ranged from 4 (very satisfied) to 1 (dissatisfied). Sowan and colleagues [25] used a 5-point Likert-type scale that ranged from 5 (very satisfied) to 1 (very unsatisfied). Both Petroulias [10] and Bowers and colleagues [24] used anecdotes and questionnaires to assess patient satisfaction with the information provided, respectively. The results of all studies indicated that patients were satisfied or very satisfied with the information and multimedia devices. However, some patients in the multimedia group showed increased anxiety, especially those with a history of complications [8]. Table 5 summarizes effect of multimedia-based education on patient satisfaction.

Table 5. Effect of multimedia-based education on patient satisfaction improvement.

Author, country	Design	Interventions	Follow-up (days)	Assessment tool	Results summary
Veyrier et al [8], France	Prospective case-control	OP[a] plus cartoon and card game versus OP alone (control)	1-2	Likert-scale	Patients in the test group were satisfied or very satisfied with the intervention and the quality of the documents provided.
Bowers et al [24], Canada	Quasi-experimental study	OP plus video presentation versus OP alone (control)	1	Questionnaire	The overall satisfaction of patients in the test group was significantly higher than that in the control group.
Petroulias [10], USA	Quasi-experimental study	OP plus video presentation versus OP alone (national data)	42	Questionnaire	Overall satisfaction with learning how to maintain the peripherally inserted central catheter at home improved after the video presentation.
Sowan et al [25], USA	Quasi-experimental study	OP plus multimedia patient decision aids versus OP alone (control)	1-2	Questionnaire	Overall patient satisfaction scores improved in a similar way between groups.

[a] OP: oral presentation.

Incidence of Catheter-Related Complications

Three studies assessed the incidence of catheter-related complications [8-10], but meta-analysis was not possible due to the different study designs used. There was a randomized controlled trial [9], a case-control study [8] and a quasi-experimental study [10]. In addition, the follow-up period varied across studies (42 days [10], 90 days [9], 10 months [8]). However, in all the studies, the incidence of catheter-related complications was slightly lower in the multimedia group than in the control group. The common complications observed in the multimedia group were infection, thrombosis, and displacement. Table 6 summarizes the effect of multimedia-based education on incidence of catheter-related complications.

Table 6. Effect of multimedia-based education on incidence of catheter-related complications.

Author, country	Design	Interventions	Follow-up (days)	Assessment tool	Results summary
Veyrier et al [8], France	Prospective case-control	OP[a] plus cartoon and card game versus OP alone (control)	305	Medical records	The risk of adverse events in the multimedia group was 38.5% lower than that in the control group.
Emery et al [9], USA	Randomized controlled trial	OP plus VOIPP[b] versus OP alone (control)	90	Questionnaire	All-cause readmissions in the VOIPP group was 14.4% lower than that in control group. The incidence of catheter-related blood stream infections in the VOIPP group was 10.6% lower than that in control group.
Petroulias [10], USA	Quasi-experimental study	OP plus video presentation versus OP alone (national data)	42	Medical records	The incidence of catheter occlusion was 0% in the video group and 14%-36% nationally (national data).

[a] OP: oral presentation.

[b] VOIPP: voiceover interactive PowerPoint.

Seeking Medical Care After the Occurrence of Complications

Two studies evaluated the effect of intervention on seeking medical care after the occurrence of complications [8,9], but meta-analysis was not possible due to different reporting methods. Emery and colleagues [9] used the number of calls received by health care providers, while Veyrier and colleagues [8] used the number of delayed visits after complications occurred. Communication between patients and medical staff in the multimedia group was improved, and there were fewer cases of delayed treatment after complications. The number of calls to health care providers in the multimedia group increased significantly compared to the control group. Patients in the multimedia group were more likely than those in the control group to report or seek clarification from their healthcare provider if problems arose [9]. Likewise, the number of cases of delayed medical treatment after complications decreased significantly in the multimedia group compared with the control group [8]. Table 7 summarizes the effect of multimedia-based education on seeking medical care after the occurrence of complications.

Table 7. Effect of multimedia-based education on seeking medical care after the occurrence of complications.

Author, country	Design	Interventions	Follow-up (days)	Assessment tool	Results summary
Veyrier et al [8], France	Prospective case-control	OP[a] plus cartoon and card game versus OP alone (control)	305	Medical records	Risk of delay to sick medical care after adverse event occur was lower in the test group compared to the control group (0% versus 100%).
Emery et al [9], USA	Randomized controlled trial	OP plus VOIPP[b] versus OP alone (control)	90	Medical records	Patients in the test group were more likely than those in the control group to report or seek clarification from their healthcare provider if problems arose.

[a] OP: oral presentation.

[b] VOIPP: voiceover interactive PowerPoint.

Discussion

Principal Findings

In recent years, there have been many suggestions to provide multimedia-based home catheter management education for patients with peripherally inserted central catheters [9,11,24].Our study determined the potential benefits of multimedia-based home catheter management education for patients with peripherally inserted central catheters and discusses the clinical implications. We identified 4 major potential benefits for patients: (1) improved knowledge, (2) increased satisfaction, (3) reduced incidence of catheter-related complications, and (4) reduced number of cases of delayed medical care after complications. The generalizability of the findings is limited due to the low methodological quality, small number of studies, and the use of multiple forms of multimedia across studies.

Multimedia-based education approach can improve home catheter management knowledge of patients living with peripherally inserted central catheter. In this review, 4 out of 5 studies reported that multimedia-based education had a beneficial impact on patients by improving their knowledge of peripherally inserted central catheter management [8,11,24,25]. Bowers and colleagues [24] found that patients in the multimedia presentation group scored significantly better on comprehension than those in the control group. Fusco and colleagues [11] found that the overall knowledge score of the video supplement group was significantly higher than that of the interview group or the interview plus brochure group. Veyrier and colleagues [8] found that patients' comprehension improved significantly when they were given two pharmaceutical consultations using cartoons and a card game demonstration, compared with a control group. Furthermore, Sowan and colleagues [25] found that patients in the multimedia patient decision aids group had higher levels of knowledge recall and retention than patients in the control group who were given information orally. This is likely because traditional peripherally inserted central catheter management education relies on patients passively acquiring knowledge through didactic sessions and instructional manuals, whereas most patients actively request to learn at their own pace and view pictures related to the information presented. Patients, especially those diagnosed with cancer, are often intellectually challenged by the diagnosis of the disease and pay little attention to the verbal information provided by medical staff. Therefore, adopting flexible learning approaches, such as multimedia-based education, can help patients to resume learning more frequently when they are in a good mood. However, Emery and colleagues [9] found no statistically significant difference in overall knowledge scores between the voiceover interactive PowerPoint supplement group and the control group immediately after the intervention or 7 to 10 days later. The variability of the results of later studies may be due to the mix of educational content without considering the impact on the outcome of the intervention. In the study by Emery and colleagues, participants were given information about the tunneled catheter and peripherally inserted central catheter, but the authors did not explain how this confounding factor was considered in the questionnaire and data analysis [9].

Multimedia-based education can improve patients' satisfaction with information and services provided. In the current review, 4 studies evaluated patient satisfaction with information and services provided [8,10,24,25], but meta-analysis was excluded due to the use of different assessment tools. Nevertheless, all studies showed that patients were satisfied with the peripherally inserted central catheter information and services. Bowers and colleagues [24] found that satisfaction with peripherally inserted central catheter information and services during the informed consent process was significantly higher in the multimedia presentation group than in the control group. Similarly, patients in the multimedia groups were more satisfied with peripherally inserted central catheter information and devices [8,10,25]. These improvements may be related to multimedia technology, which allows patients to access more dynamic and easily understood information at their fingertips, rather than the traditional didactically taught approach, which requires patients to see a health care provider. However, continued access to some information about peripherally inserted central catheter, such as images depicting the location of the catheter in the heart, has been reported to increase anxiety, especially in patients with a history of complications [8]. The results further emphasize the need to customize peripherally inserted central catheter management information based on patient preferences and needs, which is of great value in improving patient satisfaction with the information and services provided.

Multimedia-based education can moderately reduce the incidence of catheter-related complications. Three studies evaluated the effect of multimedia-based education on catheter-related complications [8-10], but meta-analysis was not possible due to different study designs. All studies reported a slight decrease in catheter-related complications incidence in the multimedia-based peripherally inserted central catheter management education group compared to the control group. This may be because in order for patients to be able to make decisions for the peripherally inserted central catheter and to accurately implement the recommendations of their healthcare provider, they need to be fully aware of catheter management skills. Compared with the traditional didactic education approach, the peripherally inserted central catheter management education based on multimedia not only improves the knowledge level of patients, but also improves their satisfaction with information and services provided. Petroulias [10] successfully reduced the occlusion rate of patients by using electronic tablets loaded with video to educate patients on how to flush peripherally inserted central catheter line. Patients received FaceTime training and watched videos showing the 10 steps of flushing. The results showed that the occlusion rate of peripherally inserted central catheters was significantly lower than when traditional teaching methods were used. Likewise, Emery and colleagues [9] found that the voiceover interactive PowerPoint group had a 10.6% lower incidence of catheter-related blood stream infections than the control group.

Multimedia-based education can reduce medical care delays when complications occur. Two studies assessed the risk of delay in seeking medical care when complications occurred [8,9], but due to different reporting methods, meta-analysis was not possible. Emery and colleagues [9] found that patients in the voiceover interactive PowerPoint group communicated with their health care provider more frequently than patients in the control group. Most calls were for information about the catheter itself, including its site problem, occlusion or rupture, and dressing problem. Similarly, Veyrier and colleagues [8] found that the strip cartoon and card game group had a lower risk of delayed medical care after an adverse event than the control group. This is likely because better communication between patients and healthcare providers not only makes it easier for patients to follow-up, but also clarifies problems they still do not understand. In addition, the improvement of patients' knowledge of peripherally inserted central catheter management is also considered to enable patients to pay more attention to or observe the changes in the catheter or themselves and report to the medical staff in a timely manner.

Strengths and Limitations

Our study has its strengths and limitations. The main strengths of our study include the use of a systematic approach to search for studies, the use of well-known electronic databases, comprehensive analysis of eligible study results, and extensive discussion of current study findings. However, the generalizability of current study findings is limited by several factors, including the small number of eligible studies, small sample size in some of the eligible studies, the low methodological quality of studies, the unreliable tools used to measure outcomes, the lack of long-term follow-up studies in particular to knowledge, and studies using different types of multimedia.

Conclusions

In this study, the potential benefits of multimedia-based home catheter management education for patients with peripherally inserted central catheter were highlighted, and the clinical implications were discussed. The findings revealed that peripherally inserted central catheter management education through multimedia-based presentation not only improved the knowledge level and satisfaction of patients, but also reduced the incidence of catheter-related complications and reduced the medical care delay when complications occurred. Due to the low methodological quality and number of included studies and the use of multiple multimedia formats, the generalizability of the findings is limited. Further review, including original studies of high methodological quality, is needed to confirm the current findings.

Acknowledgments

No funds were raised for the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PubMed search results summary.
[PDF File (Adobe PDF File), 177 KB - jmir_v22i12e17899_app1.pdf]

Multimedia Appendix 2
Principal findings summary and evidence strength of each outcome.
[PDF File (Adobe PDF File), 258 KB - jmir_v22i12e17899_app2.pdf]

References

1. Di Santo MK, Takemoto D, Nascimento R, Nascimento AM, Siqueira É, Duarte CT, et al. Peripherally inserted central venous catheters: Alternative or first choice vascular access? J Vasc Bras 2017;16(2):104-112 [FREE Full text] [doi: 10.1590/1677-5449.011516] [Medline: 29930634]
2. Sharp R, Grech C, Fielder A, Mikocka-Walus A, Cummings M, Esterman A. The patient experience of a peripherally inserted central catheter (PICC): A qualitative descriptive study. Contemp Nurse 2014;48(1):26-35. [doi: 10.5172/conu.2014.48.1.26] [Medline: 25410192]
3. Leung T, Lee C, Tai C, Liang Y, Lin C. A retrospective study on the long-term placement of peripherally inserted central catheters and the importance of nursing care and education. Cancer Nurs 2011;34(1):E25-E30. [doi: 10.1097/NCC.0b013e3181f1ad6f] [Medline: 20885304]
4. Chopra V, Flanders SA, Saint S, Woller SC, O'Grady NP, Safdar N, Michigan Appropriateness Guide for Intravenouse Catheters (MAGIC) Panel. The Michigan Appropriateness Guide for Intravenous Catheters (MAGIC): Results From a Multispecialty Panel Using the RAND/UCLA Appropriateness Method. Ann Intern Med 2015 Sep 15;163(6 Suppl):S1-40 [FREE Full text] [doi: 10.7326/M15-0744] [Medline: 26369828]
5. Cotogni P, Pittiruti M. Focus on peripherally inserted central catheters in critically ill patients. World J Crit Care Med 2014 Nov 04;3(4):80-94 [FREE Full text] [doi: 10.5492/wjccm.v3.i4.80] [Medline: 25374804]
6. Yap YS, Karapetis C, Lerose S, Iyer S, Koczwara B. Reducing the risk of peripherally inserted central catheter line complications in the oncology setting. Eur J Cancer Care (Engl) 2006 Sep;15(4):342-347. [doi: 10.1111/j.1365-2354.2006.00664.x] [Medline: 16968315]
7. Park JY. Implementing a central venous catheter self-management education program for patients with cancer. Eur J Oncol Nurs 2016 Dec;25:1-8. [doi: 10.1016/j.ejon.2016.08.010] [Medline: 27865247]
8. Veyrier M, Henry C, Decottignies A, Laouini T, Maitre S, Marthey L, et al. [Design and assessment of a outpatient living with PICC-line support program: A pilot study]. Ann Pharm Fr 2019 Jan;77(1):62-73. [doi: 10.1016/j.pharma.2018.07.001] [Medline: 30243470]
9. Emery D, Pearson A, Lopez R, Hamilton C, Albert NM. Voiceover Interactive PowerPoint Catheter Care Education for Home Parenteral Nutrition. Nutr Clin Pract 2015 Oct;30(5):714-719. [doi: 10.1177/0884533615584391] [Medline: 25998123]
10. Petroulias PL. Use of Electronic Tablets for Patient Education on Flushing Peripherally Inserted Central Catheters. J Infus Nurs 2017;40(5):298-304. [doi: 10.1097/NAN.0000000000000239] [Medline: 28885478]
11. Fusco F, Armando T, Storto S, Mussa M. [Efficacy of educational intervention for patients wearing peripherall inserted central catheter. A pilot study]. Prof Inferm 2016;69(1):17-26. [doi: 10.7429/pi.2016.691017] [Medline: 27191518]
12. Parás-Bravo P, Paz-Zulueta M, Santibañez M, Fernández-de-Las-Peñas C, Herrero-Montes M, Caso-Álvarez V, et al. Living with a peripherally inserted central catheter: the perspective of cancer outpatients-a qualitative study. Support Care Cancer 2018 Feb;26(2):441-449 [FREE Full text] [doi: 10.1007/s00520-017-3815-4] [Medline: 28707169]
13. Alpenberg S, Joelsson G, Rosengren K. Feeling Confident in Using PICC Lines. Home Health Care Management & Practice 2015 Jan 07;27(3):119-125. [doi: 10.1177/1084822314566300]
14. Molloy D, Smith L, Aitchison T. Cytotoxic chemotherapy for incurable colorectal cancer: living with a PICC-line. J Clin Nurs 2008 Sep;17(18):2398-2407. [doi: 10.1111/j.1365-2702.2008.02359.x] [Medline: 18705719]
15. Chen J, Zhao H, Xia Z, Zhang Y, Lv X, Zhou X, et al. Knowledge, attitude, and practice toward the daily management of PICC in critically ill cancer patients discharged from intensive care units. Cancer Manag Res 2018;10:1565-1574 [FREE Full text] [doi: 10.2147/CMAR.S166008] [Medline: 29950895]
16. Meyer BM. Understanding the Patient Experience of Peripherally Inserted Central Catheter-Related Deep Vein Thrombosis Using Interpretive Phenomenology. J Infus Nurs 2017;40(5):287-296. [doi: 10.1097/NAN.0000000000000238] [Medline: 28885477]
17. Simonetti G, Sommariva A, Lusignani M, Anghileri E, Ricci CB, Eoli M, et al. Prospective observational study on the complications and tolerability of a peripherally inserted central catheter (PICC) in neuro-oncological patients. Support Care Cancer 2020 Jun;28(6):2789-2795. [doi: 10.1007/s00520-019-05128-x] [Medline: 31724074]
18. Kumpf VJ, Tillman EM. Home parenteral nutrition: safe transition from hospital to home. Nutr Clin Pract 2012 Dec;27(6):749-757. [doi: 10.1177/0884533612464888] [Medline: 23090654]
19. Sawyer SM, McNeil R, McCarthy M, Orme L, Thompson K, Drew S, et al. Unmet need for healthcare services in adolescents and young adults with cancer and their parent carers. Support Care Cancer 2017 Jul;25(7):2229-2239. [doi: 10.1007/s00520-017-3630-y] [Medline: 28261754]
20. Hulin C, Hansen T, Heron L, Pughe R, Streetly M, Plate A, et al. Living with the burden of relapse in multiple myeloma from the patient and physician perspective. Leuk Res 2017 Aug;59:75-84 [FREE Full text] [doi: 10.1016/j.leukres.2017.05.019] [Medline: 28582636]
21. Kim Y, Shaffer KM, Rocha-Lima C, Milton A, Carver CS. Exploring the role of ethnicity on perceptions of cancer and physical health recovery during the first year of survivorship. J Health Psychol 2016 Jul;21(7):1331-1338. [doi: 10.1177/1359105314552304] [Medline: 25293968]
22. Okediji P, Salako O, Fatiregun O. Pattern and Predictors of Unmet Supportive Care Needs in Cancer Patients. Cureus 2017 May 09;9(5):e1234 [FREE Full text] [doi: 10.7759/cureus.1234] [Medline: 28620565]

23. Parás-Bravo P, Paz-Zulueta M, Sarabia-Lavin R, Jose Amo-Setién F, Herrero-Montes M, Olavarría-Beivíde E, et al. Complications of Peripherally Inserted Central Venous Catheters: A Retrospective Cohort Study. PLoS One 2016;11(9):e0162479 [FREE Full text] [doi: 10.1371/journal.pone.0162479] [Medline: 27588946]
24. Bowers N, Eisenberg E, Montbriand J, Jaskolka J, Roche-Nagle G. Using a multimedia presentation to improve patient understanding and satisfaction with informed consent for minimally invasive vascular procedures. Surgeon 2017 Feb;15(1):7-11. [doi: 10.1016/j.surge.2015.09.001] [Medline: 26464072]
25. Sowan AK, Beraya AR, Carrola A, Reed C. Effect of a Multimedia Patient Decision Aid to Supplement the Informed Consent Process of a Peripherally Inserted Central Venous Catheter Procedure: Pre-Post Quasi-Experimental Study. JMIR Med Inform 2018 Nov 28;6(4):e11056 [FREE Full text] [doi: 10.2196/11056] [Medline: 30487117]
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
27. Higgins J, Altman D, Sterne J. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Churchill R, Chandler J, Cumpston MS, editors. Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017). West Sussex PO19 8SQ, England: The Cochrane Collaboration; 2017.
28. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
29. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
30. Ma D, Cheng K, Ding P, Li H, Wang P. Self-management of peripherally inserted central catheters after patient discharge via the WeChat smartphone application: A systematic review and meta-analysis. PLoS One 2018;13(8):e0202326 [FREE Full text] [doi: 10.1371/journal.pone.0202326] [Medline: 30153253]

Edited by G Eysenbach; submitted 20.01.20; peer-reviewed by H Osborne, J Appalasamy; comments to author 12.06.20; revised version received 07.07.20; accepted 11.11.20; published 10.12.20.

Please cite as:
Malale K, Fu J, Nelson W, Gemuhay HM, Gan X, Mei Z
Potential Benefits of Multimedia-Based Home Catheter Management Education in Patients With Peripherally Inserted Central Catheters: Systematic Review
J Med Internet Res 2020;22(12):e17899
URL: http://www.jmir.org/2020/12/e17899/
doi:10.2196/17899
PMID:33300884

Original Paper

Patient Challenges and Needs in Comprehending Laboratory Test Results: Mixed Methods Study

Zhan Zhang[1*], PhD; Daniel Citardi[1*], MSc; Aiwen Xing[2*], MSc; Xiao Luo[3*], PhD; Yu Lu[1*], MSc; Zhe He[4*], PhD

[1]School of Computer Science and Information Systems, Pace University, New York, NY, United States
[2]Department of Statistics, Florida State University, Tallahassee, FL, United States
[3]School of Engineering and Technology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, United States
[4]School of Information, Florida State University, Tallahassee, FL, United States
[*]all authors contributed equally

Corresponding Author:
Zhan Zhang, PhD
School of Computer Science and Information Systems
Pace University
1 Pace Plaza
New York, NY, 10078
United States
Phone: 1 9147733254
Email: zzhang@pace.edu

Abstract

Background: Patients are increasingly able to access their laboratory test results via patient portals. However, merely providing access does not guarantee comprehension. Patients could experience confusion when reviewing their test results.

Objective: The aim of this study is to examine the challenges and needs of patients when comprehending laboratory test results.

Methods: We conducted a web-based survey with 203 participants and a set of semistructured interviews with 13 participants. We assessed patients' perceived challenges and needs (both informational and technological needs) when they attempted to comprehend test results, factors associated with patients' perceptions, and strategies for improving the design of patient portals to communicate laboratory test results more effectively. Descriptive and correlation analysis and thematic analysis were used to analyze the survey and interview data, respectively.

Results: Patients face a variety of challenges and confusion when reviewing laboratory test results. To better comprehend laboratory results, patients need different types of information, which are grouped into 2 categories—generic information (eg, reference range) and personalized or contextual information (eg, treatment options, prognosis, what to do or ask next). We also found that several intrinsic factors (eg, laboratory result normality, health literacy, and technology proficiency) significantly impact people's perceptions of using portals to view and interpret laboratory results. The desired enhancements of patient portals include providing timely explanations and educational resources (eg, a health encyclopedia), increasing usability and accessibility, and incorporating artificial intelligence–based technology to provide personalized recommendations.

Conclusions: Patients face significant challenges in interpreting the meaning of laboratory test results. Designers and developers of patient portals should employ user-centered approaches to improve the design of patient portals to present information in a more meaningful way.

(J Med Internet Res 2020;22(12):e18725) doi:10.2196/18725

KEYWORDS

consumer health information; health information technology; patient portal; clinical laboratory test

Introduction

Motivation

Rapid and accurate communication of laboratory test results to patients is critical [1]. Growing evidence suggests that patients are increasingly interested in direct access to their test results, regardless of the indication of the results [2,3]. Policies, guidelines, and initiatives have been implemented to encourage health care organizations to provide patients with easy access to their test results through patient-facing technologies [4,5], such as patient portals, which are often linked to electronic health record systems [6,7]. It is well recognized by the health informatics community that increasing patients' access to their data can lead to better patient-centered medical care [8], greater patient engagement in medical decision making [3,4,9], and enhanced patient-clinician relationship [10,11]. Despite the potential benefits of such direct access, there are long-standing yet unresolved concerns that many patients may not be able to fully understand their laboratory test results and evaluate the indications for their health [12,13].

Why does merely providing access to laboratory test results not guarantee comprehension? First, laboratory tests are complex. Even medical professionals cannot efficiently and accurately interpret them [14], let alone patients. Second, many patients have limited health literacy and numeracy, thereby making it difficult to interpret the test results and decide corresponding actions (eg, call their doctor for an urgent appointment vs regular follow-up visits) [15,16]. Finally, the current interface of patient portals also poses significant challenges in viewing and comprehending test results [15,17,18]. Many portals only provide test results to patients in a tabular format with standard reference ranges, similar to the format received by clinicians. As such, these portals could be of little use to those who are supposed to benefit from them most [15,19]. Although some visual aids, such as color, have been used in patient portals to help patients identify out-of-range values [20], test results are often presented with unfamiliar abbreviations and units, lacking guidance on whether the results are concerning [21]. Therefore, lay individuals have difficulty identifying meaningful information out of their test results and making informed decisions [22].

Despite the urgent need to address these issues, only a few studies have looked into how to better deliver test results along with the support of people with specific needs [23,24]. There is a lack of in-depth, empirical understanding of patients' experiences, challenges, and needs related to comprehending laboratory test results via patient portals. To that end, we conducted a mixed methods study to address this research gap by examining the 3 following research questions (RQs):

- RQ1: What are patients' perceived challenges and needs when they attempt to comprehend test results?
- RQ2: What factors are associated with patients' perceptions?
- RQ3: How can we improve the design of patient portals to communicate laboratory test results to patients in a more meaningful and understandable manner?

Related Work

The literature on supporting patients' comprehension of test results is sparse but growing. A few studies attempted to understand patients' confusion about their laboratory test results by examining online posts. For example, Reynolds et al [25] examined patient question posts in an online health forum and found that patients asked questions about diagnosis, treatment, laboratory report, risk, and prognosis. In a similar vein, Zhang et al [26] examined question posts in a social question and answer site and found that patients' confusions primarily centered on understanding the meaning of laboratory values, specific terminology, and the causes of abnormal or inconsistent results. However, one major limitation of these studies is the lack of direct patient input; as indicated by previous work, failing to involve patients in the process of developing patient-facing applications could lead to limited uptake of these systems [27]. Therefore, a comprehensive understanding of patients' needs is of utmost importance for designing supportive technologies.

A few studies addressed this limitation by eliciting patient opinions and insights through user studies, such as surveys and interviews. For example, Giardina et al [23] examined patient perceptions of receiving test results via online portals and found that most patients did not receive any explanatory information at the time they received the result. They also stated that current online portals are not designed to present test results to patients in a meaningful way. In another study, Reiff et al [28] conducted semistructured interviews with parents of pediatric outpatients to identify the needs of families undergoing chromosomal microarray (CMA) testing. They found that incomplete comprehension of test results and scientific uncertainty were prominent challenges faced by families. Receiving results from nongeneticists and misleading internet searches, among others, was found to contribute to the misunderstanding of CMA testing results. Building upon their work, we conducted a mixed methods study to gain further insight into patients' challenges and needs when comprehending laboratory test results.

Gaining an in-depth and thorough understanding of patients' information needs, faced challenges, and preferences when comprehending laboratory test results is vital to inform the design of patient portals and other patient-facing informatics tools [21]. However, only a few studies explored patients' information needs and challenges when interpreting their test results and how to improve the design of patient portals for better support. For example, Nystrom et al [29] designed and evaluated a patient-centered test result interface that consists of visual ranges of laboratory values, nontechnical descriptions of the test and the result, and links to reputable internet resources where patients could learn more about their result. User testing indicates that these features were perceived as usable because they account for patients' needs and preferences for understanding laboratory test results. Zikmund-Fisher et al [21] tested the effect of presenting patients with an additional reference point in visual displays of laboratory test results regarding when test results become clinically concerning. They found that providing patients with such evaluative cues can substantially reduce the perceived urgency of out-of-range results that normally do not require immediate action. These studies, however, did not have a user study component; therefore, patients were left out of the system design process. This study contributes to this line of research by examining patients' needs through holistic user studies and leveraging the findings to suggest design implications for patient portals to promote patients' comprehension of test results.

Aim of This Study

The ultimate goal of this study is to understand the design requirements and specific strategies for improving patient portals to communicate laboratory test results to patients in a way that is more meaningful and understandable. The development of these requirements and strategies requires an empirical understanding of patients' perceived challenges and needs when attempting to comprehend test results.

Methods

Design

We used a mixed methods approach [30], combining a web-based survey, and semistructured interviews, to gain an empirical and thorough understanding of patients' experiences, challenges, and needs in comprehending laboratory test results. The reason for using both survey and interview methodologies is that they are suitable for creative knowledge generation of a multilayered issue [31], such as our research context, which encompasses many interdependent factors (eg, people, technology, data, and knowledge). In addition, these 2 methods are complementary to each other, allowing us to compare and triangulate the findings from each study [32]. For example, the survey study allowed us to broadly understand our research questions, whereas the interview study helped us obtain more in-depth insights. Both methods were used jointly to ensure a mix of both broad and deep understanding of the 3 research questions. The survey (Multimedia Appendix 1) was developed in an iterative manner by the researchers and pilot-tested with a small group of people (n=10) to ensure the clarity, appropriateness, and relevance of the questions. The interviews (Multimedia Appendix 2) were informed by the exploratory survey. This study was approved by the Pace University Institutional Review Board (IRB# 19-08).

Data Collection

Data collection occurred between July and September 2019. We recruited 203 participants from Amazon Mechanical Turk (MTurk) to participate in our survey. Using MTurk to study patients' perceptions has been proven reliable and effective [33]. To determine their eligibility, we first asked potential respondents whether they had recently taken any laboratory test and whether they had used patient portals to view the results. If they responded with *yes* to both questions, we asked them to read and electronically sign the informed consent form and complete the survey. Toward the end of the survey, we invited them to participate in a follow-up interview study. Potential participants were instructed to email researchers to arrange a suitable time for the interview.

Online Survey

Table 1 shows survey participant characteristics, including an almost equal number of male and female participants (104/203, 51.2% and 98/203, 48.3%, respectively). Most of the participants were White (141/203, 69.5%), aged between 26 and 49 years (129/203, 63.5%), had a bachelor's or higher degree (134/203, 66.1%), had a full-time job (157/203, 77.3%), and self-reported having above medium health literacy (111/203, 54.7%) and technology proficiency (148/203, 72.9%).

Table 1. Characteristics of survey participants (N=203).

Participant characteristics	Participants, n (%)
Age (years)	
18-25	38 (18.7)
26-49	129 (63.5)
50-64	32 (15.8)
65 and older	4 (2.0)
Gender	
Male	104 (51.2)
Female	98 (48.3)
Other	1 (0.5)
Race or ethnicity	
Asian or Pacific Islander	9 (4.4)
African American	34 (16.7)
Hispanic/Latino	12 (5.9)
American Indian	4 (2.0)
White	141 (69.5)
Other	3 (1.5)
Education	
Doctorate degree	6 (3.0)
Master’s degree	32 (15.8)
Bachelor’s degree	96 (47.3)
Associate degree	28 (13.8)
High school degree	40 (19.7)
Other	1 (0.4)
Employment status	
Unemployed	17 (8.4)
Part time	23 (11.3)
Full time	157 (77.3)
Other	6 (3.0)
Employment industry	
Government	16 (7.9)
Health care	23 (11.3)
Education	32 (15.8)
Finance	22 (10.8)
Information technology	44 (21.7)
Other	66 (32.5)
Health literacy	
High	44 (21.7)
Medium-high	67 (33.0)
Medium	81 (39.9)
Low-medium	10 (5.0)
Low	1 (0.4)
Technology proficiency	

Participant characteristics	Participants, n (%)
High	67 (33.0)
Medium-high	81 (39.9)
Medium	52 (25.6)
Low-medium	2 (1.0)
Low	1 (0.5)

The survey assessed 3 domain areas: (1) participants' sociodemographic characteristics, (2) participants' experiences and information needs about understanding laboratory results, and (3) participants' perceptions of reviewing results via patient portals. Participants' characteristics collected included age, gender, race/ethnicity, education, and occupation. As health literacy and technology proficiency may affect patient use of the patient portal to view test results [15,19,34], we also asked them to rate their technology proficiency and health literacy on a scale of 1 to 5 (*1* denotes low literacy, whereas *5* denotes high literacy). Patients' experiences include whether they understand the results, what actions have been taken to make sense of the test results, if they have confusion regarding the result, and what types of information or advice are needed. Patients' perceptions of the patient portal include the usability of patient portals, challenges or concerns (if any) with using patient portals, and tailoring the patient portal to their needs.

We directed participants to answer survey questions based on their most recent experience of reviewing laboratory test results in a patient portal. However, respondents may have difficulty recalling such an experience, and thus, we provided a commonly seen laboratory test—the lipid profile (eg, total cholesterol, low-density cholesterol, and high-density cholesterol)—to help participants situate themselves in the context (Table 2). The test results were presented following the format currently implemented in many patient portals, that is, using tables to show test values and units. The reason for choosing this laboratory test is that it is routinely performed among adults for screening and diagnostic purposes; therefore, many adults are likely familiar with the test [35]. The survey took about 10-15 min to complete, and the respondents received US $2 after they completed the survey.

Table 2. Lipid profile.

Type	Reference range	Results
HDL[a] cholesterol	>39 mg/Dl	52 mg/dL
LDL[b] cholesterol	0-99 mg/dL	115 mg/dL
Total cholesterol	100-199 mg/dL	185 mg/dL
Triglycerides	0-149 mg/dL	164 mg/dL

[a] HDL: high-density lipoprotein.

[b] LDL: low-density lipoprotein.

Semistructured Interview

To further understand patients' information needs and challenges related to understanding laboratory test results, we conducted semistructured interviews with 13 people who had participated in the previous survey (Table 3). The interviews were conducted via Skype by 2 trained researchers and lasted from 30 min to 1 hour. During the interview, we asked participants to walk through the process of receiving a test result and probed them with questions about their understanding of the results, information needs, challenges faced, perceptions of patient-provider communication, and facilitators and barriers to using the patient portal. All the interviews were audio recorded with the participants' permission and transcribed verbatim. All the participants received a US $20 Amazon gift card as compensation for their time.

Table 3. Characteristics of interview participants.

ID	Gender	Age (years)	Ethnicity	Occupation	Technology proficiency level	Health literacy level
Patient 1	Female	50-64	White	Unemployed	4	5
Patient 2	Male	50-64	African American	Retired	5	4
Patient 3	Female	50-64	White	Retired	5	5
Patient 4	Male	50-64	White	Technical analyst	5	5
Patient 5	Male	26-49	White	IT clerk	5	3
Patient 6	Male	26-49	White	Dog trainer	4	3
Patient 7	Female	50-64	White	Landscaper	3	4
Patient 8	Female	50-64	White	Executive assistant	5	5
Patient 9	Male	18-25	Asian	Medical assistant	4	4
Patient 10	Male	26-49	Asian	Visual artist	5	5
Patient 11	Female	26-49	White	Office assistant	2	4
Patient 12	Female	26-49	White	Sales	5	5
Patient 13	Female	26-49	White	Unemployed	4	4

Data Analysis

We used descriptive statistics to analyze the survey data. Correlations between the variables were evaluated using the chi-square test or Fisher exact test where appropriate. In particular, we tested the association between several intrinsic factors (eg, age, gender, educational background, technology proficiency, health literacy, and normality of results) and participants' perceptions of using patient portals to review laboratory results. All the tests were 2-sided, and the level of significance was set at 0.05. Statistical analysis was conducted using IBM SPSS 25.0. Detailed quantitative results are provided in Tables 4 and 5 and in the tables in Multimedia Appendix 3.

Table 4. Patient perceptions of interpreting test results and the association to normality of laboratory test results and health literacy (n=199).

Patient perception	Total[a], n (%)	Normality of results (*P* value)	Health literacy (*P* value)
Did your physician communicate with you about the test results before you viewed the results?			
Yes	112 (56.3)	.001[b]	.18
No	87 (43.7)	.001[b]	.18
Did you understand the result?			
Yes	120 (60.3)	.12	.03[c]
No	9 (4.52)	.12	.03[c]
Not sure	70 (35.2)	.12	.03[c]
How did you know the test result was abnormal or normal?[d]			
Visual cue on patient portals	86 (43.2)	.98	.02[c]
Clinician's explanation	104 (52.3)	.60	.21
Personal knowledge	70 (35.2)	.90	.50
Other	11 (5.5)	.59	.18
What kinds of confusion did you have?[d] (n=136)			
Normal versus abnormal laboratory result	43 (31.6)	.84	.64
Medical terminology	83 (61.0)	.37	.01[c]
Meaning of the results	54 (39.7)	.66	.10
Effects on my future health	47 (34.6)	.80	.61
Treatment option	27 (19.9)	.69	.68
Other	3 (2.2)	.87	.16
How did you feel when you saw the result?			
Negative	29 (14.6)	<.001[e]	.001[b]
Positive	144 (72.4)	<.001[e]	.001[b]
Indifferent	26 (13.1)	<.001[e]	.001[b]
What actions have you taken after viewing your test results?[d]			
Spoke with family and/or friends	101 (50.8)	.86	.45
Looked up information on the web	123 (61.8)	.79	.60
Posted questions in online health forums	22 (11.1)	<.001[e]	.89
Created graph of results	9 (4.5)	.30	.41
Emailed doctor	19 (9.6)	.19	.27
Called doctor	22 (11.1)	.77	.88
Made doctor's appointment	34 (17.1)	.09	.09
Other	11 (5.5)	.89	.27
Do you need more information to interpret test results?			
Agree	118 (59.3)	.005[b]	.01[c]
Neither agree nor disagree	58 (29.2)	.005[b]	.01[c]
Disagree	23 (11.6)	.005[b]	.01[c]
What types of information do you need?[d] (n=119)			
Prognosis	55 (46.2)	.007[b]	.68

Patient perception	Total[a], n (%)	Normality of results (*P* value)	Health literacy (*P* value)
Treatment options	79 (66.4)	.86	.11
What to do or ask	57 (47.9)	.50	.74
Lifestyle changes	42 (35.3)	.50	.72
Connecting with local support group	4 (3.4)	.57	.48
Other	1 (0.8)	.45	.27

[a] In total, 4 participants could not remember the normality of their test results. These 4 participants were excluded from this analysis; therefore, the total number of cases was 199.

[b] Results with *P* value<.01.

[c] Results with *P* value<.05.

[d] Participants can select more than one option for those questions.

[e] Results with *P* value<.001.

Table 5. The association between 2 factors (health literacy and technology proficiency) and patient perceptions of using patient portals to understand and review laboratory test results (N=203).

Patient perceptions	Participants, n (%)	Health literacy (*P* value)	Technology proficiency (*P* value)
I am comfortable with using patient portals to review my laboratory results			
Agree	185 (91.1)	.06	.09
Neither agree nor disagree	18 (8.9)	.06	.09
Disagree	0 (0.0)	.06	.09
I never had any trouble checking my test results on the patient portal			
Agree	161 (79.3)	.38	.002[a]
Neither agree nor disagree	25 (12.3)	.38	.002[a]
Disagree	17 (8.4)	.38	.002[a]
I find the patient portal can make me review my tests quickly			
Agree	187 (91.6)	.53	<.001[b]
Neither agree nor disagree	15 (7.4)	.53	<.001[b]
Disagree	1 (1.0)	.53	<.001[b]
I find the patient portal is useful to understand my laboratory results			
Agree	166 (81.3)	.15	.02[c]
Neither agree nor disagree	30 (14.8)	.15	.02[c]
Disagree	7 (3.9)	.15	.02[c]
I have used the resources provided by patient portals to understand my results			
Agree	131 (64.0)	.03[c]	.22
Neither agree nor disagree	40 (19.7)	.03[c]	.22
Disagree	32 (16.3)	.03[c]	.22
Is there anything that would make the portal better for you?[d]			
Make it more user-friendly	61 (29.1)	.70	.82
Allow me to send a message to my physician	83 (40.0)	.04[c]	.009[a]
Include a health encyclopedia that contains more information about the test	99 (47.3)	.48	.32
Provide timely test result explanation and follow-up instructions	108 (52.7)	.007[a]	.07
Other	2 (1.0)	.03[c]	.53

[a] Results with *P* value *P*<.01.

[b] Results with *P* value *P*<.001.

[c] Results with *P* value *P*<.05.

[d] Participants can select more than one option for those questions.

We used thematic analysis [36] to analyze the interview data. We first reviewed the interview transcripts to obtain an overview of the context. In the subsequent stage, we transferred data into NVivo (QSR International), a program for organizing, storing, and manipulating qualitative data. Two authors conducted open coding over the interview data and met regularly to discuss codes and code definitions, consolidate and refine codes until they reached consensus. In the second round of analysis, coded data were grouped under themes using affinity diagrams [37]. Themes describing the challenges and information needs of interpreting laboratory results and perceptions of receiving and reviewing results via current delivery mechanisms emerged after the second round of coding.

Results

Challenges in Comprehending Laboratory Test Results

The results showed that only 55.2% (112/203) of participants reported that they were contacted by their physician before reviewing the results in the patient portal, suggesting that it is not always the case for patients to receive timely notification and explanation of the results from their physician. One participant explained: “I didn’t actually talk to the doctor until probably like a week after that, because I wasn't able to get a hold of them” (P10). This finding is consistent with previous work stating that patients usually did not receive any explanatory information from their physician when they received the result [38]. Compared with patients who received normal results, those with abnormal results were more likely to be contacted (*P*=.001). When the physician explanations were absent, it is surprising to see that some participants chose not to bother their physician:

> *We have these mega practices, they’re incredibly busy, and I think if I did, I’d probably be afraid of being labeled as a trouble patient, and you don't want to be a trouble patient. I want to reserve those chips for when I really need them.* [P3]

We also found that participants with relatively high technology proficiency were more likely to possess positive opinions regarding the usefulness of patient portals (eg, the patient portal was useful in reviewing the test quickly [187/203, 91.1%; *P*<.001] and understanding their laboratory results [166/203, 81.3%; *P*=.02]). In a similar vein, participants with relatively high health literacy would be more likely to use resources provided by patient portals to make sense of the laboratory results (131/203, 64%; *P*=.03). These findings align with previous research showing that there is a digital divide in the use and adoption of health technologies—people with higher literacy and numeracy can benefit more from using health technologies to manage their health [19].

Finally, our participants indicated that existing patient portals lack sufficient and useful information for patients to understand their results: “In my particular case, I am trying to understand the meaning of the various test results. I didn't see anything on the patient portal that helped me with that at all” (P4). The lack of timely explanation from a primary care physician exacerbates this challenge, leaving patients entangled with how to interpret the meaning of their results accurately.

Patients’ Confusion About Laboratory Test Results

Given these challenges, it is not surprising that many participants have difficulties understanding their laboratory results. We found that 70 survey participants (70/199, 35.2%) were unsure whether they understood their laboratory results. We believe this is because making sense of laboratory test results is a multilayered issue. Patients need to not only know whether they received a normal or abnormal result but also understand the meaning and indications of the test results. Despite participants reporting using different mechanisms to interpret the normality of test results (eg, visual cues on patient portals and personal knowledge), we found that 136 out of 203 participants (136/203, 67.0%) still reported having confusion about different aspects of their laboratory results, including medical terminology, reference range, the meaning of laboratory value, and the effects on their health care.

Participants deemed technical jargons, such as medical terminology and abbreviations, to be most confusing (83/136, 61.0%):

> *Just some of the specifics to the result. Like in the blood tests, words like neutrophils, or CBC, or bilirubin, however you pronounce that. Those very specific things that as a person who is not trained in medicine, I didn't understand without having to consult the internet.* [P4]

Although some patient portals have started implementing consumer-friendly vocabularies [39], this challenge persists—those explanations of medical terms may use additional medical jargon, which further confuses the patients. This perception was echoed by other participants who stated that patient portals overly use professional medical terms that are not easy to understand for lay individuals.

Our participants also had trouble differentiating between normal and abnormal laboratory results (43/136, 31.6%), in particular, whether a specific laboratory value falls into the normal range: “Some of the numbers don't even have a point of reference for me, so I don't understand that” (P12). In addition, some participants also reported confusion about some test results that are not in a numerical format, such as *false positive* and *nonreactive*. One participant explained:

> *I get what the HIV means. And then like you actually even click further to details and it didn't even say “Negative”, it is actually “Reactive”. That’s what I didn't understand. I didn't know what non-reactive meant.* [P6]

For those who received out-of-range test results, they reported that they struggled to understand how bad their abnormal results were and how alarmed they should feel.

Finally, participants expressed that they lacked an understanding of the meaning of the results (54/136, 39.7%) and the effects on their overall health care (47/136, 35.6%):

> *I don't know exactly what it (test result) meant clinically and for my overall health. I just kind of had a general idea. So I’m undergoing cancer treatment and so they’re checking my white blood cell counts. Those were low, but I don't really understand are they too low? I don't know necessarily what all of the things are. Like the neutrophils and I don't know what all of that means if it’s out of range.* [P13]

Unmet Information Needs

A considerable number of participants (119/203, 58.6%) expressed an interest in obtaining more pertinent information in addition to that provided by patient portals to interpret their test results. In particular, both normality of the laboratory test (*P*=.005) and health literacy (*P*=.01) were significantly associated with this requestparticipants with abnormal results (43/58, 74.1%) and relatively high health literacy were more interested in receiving additional information (116/192, 60.4%). Some participants would like to know more generic information, such as reference range and diagnostic abilities of a specific test (eg, “I guess I liked to know an explanation of what the normal levels should be [...] and what each of the tests was testing” [P9]). In contrast, others preferred personalized and contextual information (eg, “If it was more personalized based on a patient’s history, that would definitely be more helpful because then it would be something you could really know [if you] need to act on this or not" [P12]). The types of most needed personalized information include treatment options, what to do or ask, prognosis, and lifestyle changes.

More specifically, many of the participants (79/119, 66.4%) would like to know more information about treatment options, including medication and medical procedures: "I want to know something like treatment options or medications" (P8). Of particular interest here is that 35.5% (50/119) of the participants whose results were normal also expressed an interest in knowing the treatment options. One possible explanation is that the participants were asked to report the normality of their most recent laboratory test result. Therefore, some participants may receive normal results for the latest test, but they may have received an abnormal result before. We speculate that these participants expressed their general preferences and information needs not only based on the most recent result but also on their overall experience.

Our participants also needed assistance in deciphering the meaning of the results and understanding the potential prognosis (56/119, 47.1%):

> *I would like to see a little more information as far as what might be causing it if there is a result that's not normal. I would like to see a little more as far as what possibly could be there.* [P12]

Interestingly, laboratory test normality was strongly associated with whether participants would like to know prognosis information (P=.007).

There was also a demand to be informed about what they could do to cope with the bothersome symptoms they were experiencing or what they should ask the doctor during the next clinic visit (58/119, 48.7%): "I liked to know suggestions of what I should do. […] If I get an abnormal result, I would be wondering what I could be doing to improve" [P9]. In particular, people also wanted to know whether changing their lifestyle (eg, diet and exercise) could be of any help (42/119, 35.3%). Knowing such information could help them better manage their health care (eg, take actions to address elevated creatinine levels) and make informed decisions (eg, whether and when to see a doctor).

Emotional Aspect of Viewing Laboratory Results

Despite only 29 out of 203 participants (29/203, 14.3%) reporting the experience of negative emotions (eg, concerns, anxiety, and frustration) while viewing the results, we noticed that laboratory result normality was strongly associated with feelings (P<.001). This is consistent with previous work [23]—participants with abnormal results were more likely to experience negative emotions (25/58, 43.1%). One participant shared her story and expressed the necessity of getting emotional support when the result is abnormal:

> *I was diagnosed having cancer. So I need to do a bunch of lab tests regularly to keep track of everything. I was kind of scared every single time when I was checking the test results online. I hoped to see improved numbers. But they were not always good and when that happened, my level of anxiety skyrocketed in a second. It can be very tough.* [P10]

Actions Taken

To fulfill their knowledge gap and emotional needs, participants took several actions, including speaking with family and/or friends (101/203, 49.6%), looking up information online (124/203, 61.1%), posting questions in online health care forums (23/203, 11.3%), emailing and/or calling a doctor (42/203, 20.7%), and making a doctor's appointment (35/203, 17.2%). These findings show that many people sought information about their results from sources other than their physician [23].

In particular, laboratory test normality was significantly associated with posting questions online (P<.001)—participants with abnormal results would be more likely to post questions in online health care forums compared with those receiving normal results (14/58, 24% vs 8/141, 5.7%). By doing so, they were able to prepare questions or sought relevant information before their next clinic visit. However, several participants reported finding misleading and/or disturbing information on the internet (eg, "once you hit the forums, you often get crazy time"). A few participants also mentioned that online searches helped clarify medical terms but did not yield much useful information for comprehending specific results (eg, "It's too general. It's not helpful").

Strategies to Improve Patient Portals

When asked how to make the patient portal better and useful for the comprehension of laboratory results, participants ranked "provide timely test result explanation and follow-up instructions" (108/203, 52.7%), "include a health encyclopedia that contains more information about the test" (99/203, 47.3%), and "allow me to send a message to my physician" (83/203, 40.8%) as the top 3 needed features.

Our participants also emphasized the importance of making the patient portals "more user-friendly" (61/203, 29.1%). For instance, patient portals should be designed with patients in mind rather than for clinician interpretation only, as one participant stated:

> *So maybe making it less like I'm looking at the system that the doctor would be using and more something that's designed with the patient in mind and understanding who the patient is and not somebody who's going to use like the very specific language.* [P6]

They further suggested that visualizing historical results ("What I really, really think needs to be done is the ability to track your abnormal results over time") and using lay terms to explain the result ("make it like a 10-year-old would understand") could be very useful in improving the user experience of patient portals.

The design of patient portals should also take into consideration marginalized user groups, including patients with disability and older adult patients, to minimize the disparities in the use of patient portals [40]. One participant explained why this issue is important to address:

> *I noticed with patient portals is they're not very accessible to people with vision issues. It doesn't have to be fancy, but a little button that says "please push F to read this for them", and then they could hear it. [...] or maybe the doctor gives a specialized URL and he like "hey, go to this URL, and it will be for blind assist" [...] I would say it needs to be a little bit more use-friendly as far as older folks go. [...] So since the world is having a national aging problem, it should be geared more for grandma to understand it.* [P5]

Finally, it is interesting to see that several participants described the potential of incorporating artificial intelligence (AI) technology into patient portals so that their data can be used to generate more personalized medical information:

> *I know right now a big part of medicine is going to be AI where you actually interact with artificial intelligent agents. They'll probably have your information right there and it'll probably make it a lot easier for you to get a lot of your answers because they'll be accessing your information right there and giving you the answers to your questions.* [P2]

Discussion

Principal Findings

Although health care providers are expanding patients' access to laboratory test results through patient portals, less than one-third of Americans had accessed this information online [7,23]. We conducted this study to better understand lay people's perceptions about reviewing and comprehending laboratory test results through patient portals, including their perceived challenges and confusions as well as informational, emotional, and technological needs. The challenges faced by patients when reviewing laboratory test results (eg, lack of explanatory information from physicians and lack of useful information on patient portals) inevitably caused confusion about different aspects of their laboratory report, such as the meaning of laboratory values and the indications on their health care. To cope with these challenges and confusion, our participants expressed the urgent need to obtain a variety of information to comprehend laboratory results, including treatment options, what to do or ask, prognosis, and lifestyle changes. The desired enhancements of patient portals include providing timely explanations and educational resources (eg, a health encyclopedia), increasing usability and accessibility, and incorporating AI-based technology to provide personalized recommendations. Finally, we found that several intrinsic factors (eg, laboratory result normality, health literacy, and technology proficiency) had significant impacts on people's perceptions of using portals to view and interpret laboratory results. For example, participants with relatively high health literacy were more likely to understand the meaning of test results and use resources provided by patient portals to make sense of their data, while people with relatively high technology proficiency tended to agree that the patient portal was easy to use and very helpful in reviewing and understanding their laboratory values. These findings highlight significant disparities in laboratory result comprehension and patient portal use among certain groups of the population.

Design Implications

At present, the design of test results in patient portals seems to assume that patients have sufficient medical knowledge about their test results. In addition, patients often did not receive explanatory information at the time they received the result. One primary reason is that many physicians would prefer explaining results during a face-to-face clinical encounter to avoid any potential miscommunications [41]. However, on the other hand, patients are increasingly interested in accessing and interpreting their results. Many of them would search online to conduct their research [23,25,26,28]. Our participants articulated that internet searches were helpful in clarifying medical terms; however, the online information was not always reliable, and they even found contradicting or disturbing information on the internet, which caused further confusion and anxiety [28]. It is therefore crucial to provide more useful, credible, and actionable information to aid patient understanding, something current patient portals do not provide [23].

With regard to the types of information to provide, our results indicate that people have different information needs, including generic information (eg, reference range and diagnostic abilities of a specific test) and personalized or contextual information (eg, treatment options, prognosis, and what to do or ask next). This variety of information needs may be due to different levels of health literacy and numeracy that people have and whether patients receive normal or abnormal results [19]. For example, some patients may not be literate enough to understand the medical terminologies or normal ranges of a test, whereas others need help with more complicated issues, such as interpreting the laboratory results in the context of their medical history. It is, therefore, important to assess the patient's knowledge level and provide corresponding support based on patients' characteristics [42,43]. One way to accomplish this is to allow patients to add their own notes about their background (eg, health literacy, medical condition, etc) to a specific section in their patient portals to enable automatic assessment and customized support. It may also be useful to provide links to consumer-friendly and trusted information sources (eg, entries in MedlinePlus) to assist those who do not have high levels of health literacy and numeracy to read and interpret laboratory content [42]. We also noticed that people with abnormal results were more likely to ask for prognosis information. This indicates that providing interpretation and other advanced information (eg, prognosis, treatment options) to patients who received abnormal results at the time of portal release should be considered a best practice [15], as it can support patients' information needs and reduce any anxiety they may experience.

We are not arguing that we should provide all kinds of information and support to patients at one time because they may overwhelm patients and create even more confusion. Instead, we believe that the delivery of additional information should be tailored to patients' individual preferences, that is, patients should have options to decide whether they want to receive additional information in patient portals and what the information is at the time of ordering the test. Those who do not want to receive additional information can opt out and wait until their physician to provide explanations. This practice aligns with the SPIKES guideline (S: setting, P: perception, I: invitation, K: knowledge, E: empathy, S: summarize) of delivering health information to patients [44]. We also see design examples where patients can click buttons or hyperlinked texts to see additional information [29]. Presenting additional information in *pop-ups* is expected to not only ease the navigation of the interface but also empower patients to decide whether they would like to see further details.

A lot of the challenges in result comprehension faced by patients can, in part, be attributed to the interface design of patient portals. We examined how patient portals can be improved to support understanding. In line with previous work [29], we conclude that merely providing access to abundant data is insufficient; instead, the patient portal should employ user-centered design strategies to help patients better interpret and manage their test results. For example, the test results should be displayed to patients in ways they can comprehend and grasp the meaning of each test result to inform their decision making and subsequent actions. For example, it would be much easier for patients to interpret their results by delivering intuitive *gist* messages (eg, “your result was normal,” “your cholesterol levels were optimal”) when presenting numerical values [15,35]. Furthermore, as our participants described, they struggled to map a particular abnormal test result to its clinical meaning. Although most patient portals provide patients with a standard reference range to discriminate between normal and abnormal test results, patients cannot necessarily understand how bad the result is and if any urgent action is needed. The literature has suggested that it might be useful to provide an additional reference point to indicate how far outside the standard range that values become clinically concerning [21]. By doing so, it will help patients better distinguish different types of abnormal test results, that is, slightly out-of-range results versus extreme values, and in turn reduce the perceived urgency of abnormal results that do not require immediate clinical action [21].

Furthermore, our participants discussed the possibility of incorporating AI technology into patient portals to make use of their clinical data to generate more personalized medical information. Indeed, with the recent rise in the capabilities of AI, it is expected that AI-empowered health care systems can intelligently search, retrieve, and present pertinent information to patients within the context of their medical history [45]. For example, although many hospitals and physicians are still using the 2.5 or 3.0 mU/L cutoff (2.5 mU/L in the first trimester and 3.0 mU/L in the second and third trimesters) for the thyroid-stimulating hormone during pregnancy, a recent study pointed out that these cutoffs are too low and may lead to overdiagnosis or even overtreatment [46]. The AI-driven patient portals should be able to recommend such relevant and up-to-date medical evidence to patients based on their test results. Patient portals may embody AI technology to change the landscape of how people receive laboratory results and seek medical advice.

Finally, although our participants wanted to be informed about the meaning of diagnostic results, needless anxiety and negative emotions may arise when communicating *abnormal* diagnostic results without the presence of a medical professional. In particular, we found that patients with abnormal results were more likely to experience negative emotions. Therefore, it is necessary to provide additional support or resources for patients to mitigate emotional stress when communicating sensitive results through patient portals. For example, patients should have options to determine whether they want to receive concerning results via the portal. This could ensure that patient portals deliver potentially stressful health information in a more empathic manner [47]. In addition, previous work has shown that patient navigators can reduce negative emotions, such as, anxiety, for patients who receive abnormal mammography while waiting for follow-up testing [48].

However, implementing these design suggestions is challenging. This is because current patient portals lack consistency in the design for communicating laboratory test results across different portal vendors, which not only frustrates patients but also creates barriers for incorporating *best practices* and new features to different portals [23,49]. Therefore, an urgent research agenda is determining how to develop standards and guidelines and encourage portal providers to incorporate them to enhance patients’ understanding of their test results.

In brief, when communicating test results via portals, it is necessary to include information about the test itself (eg, test capability, the purpose of test), easy-to-understand explanations of medical jargon, the results in the context of the patient’s health, directions for next steps, and specific educational resources [35,50,51]. Furthermore, designers and developers of patient portals should embody user-centered approaches to significantly improve the design of patient portals to present information in a more meaningful way [29,52]. Finally, national test result notification and interface design standards for patient portals should be developed to ensure interoperability and consistency in features across portal vendors. Health policies should be enacted to support these strategies [23].

Limitations and Future Research

A few limitations should be noted. First, our survey and interview participants self-reported their experiences with using patient portals to view laboratory results. Some of our respondents may have difficulties recalling such experiences. As described in the Methods section, we provided a commonly seen laboratory test to participants to cope with this issue. Second, our results may not generalize to all types of patients because we recruited our participants online and our sample inevitably consisted of a large majority of participants who self-reported to have medium to high levels of technology proficiency and health literacy. This limitation of participant recruitment could affect the generalizability of the results because there is a lack of a great representative of marginalized populations who can benefit most from this research. In our future work, we will include more marginalized populations (eg, less literate people, older adults) and examine how to improve their understanding of laboratory results. Third, this study did not investigate which patient portal our participants used. However, we believe it has a limited impact on our results as many patient portals share similar characteristics, such as displaying results in a tabular format with a reference range.

Finally, we will design and prototype a new interface of patient portals for test results based on our findings. We will involve different types of patients with various characteristics and use a multiphased, user-centered approach combined with rapid prototyping and formative evaluation.

Conclusions

In conclusion, our findings suggest that there are challenges for patients to comprehend their laboratory test results through the patient portal. This is mainly because of the lack of communication with physicians and the lack of support and useful information in current patient portals. Several factors, including participants' health literacy and technology proficiency as well as laboratory result normality, have impacts on people's perceptions of using patient portals to understand the laboratory results. Our participants have a variety of information needs, ranging from general information (eg, reference range and diagnostic abilities of a specific test) to personalized information (eg, treatment options, prognosis, and what to do or ask next). Participants also emphasized the importance of improving patient portals to better meet their needs. This study contributes to an empirical, in-depth understanding of the challenges and needs of patients in comprehending laboratory test results and informs strategies and design implications for informatics tools to promote patients' comprehension of test results.

Acknowledgments

This study was supported by the Pace University Scholarly Research Award, the National Institute on Aging of the National Institutes of Health (NIH) under award number R21AG061431 and the Florida State University-University of Florida Clinical and Translational Science Award funded by the National Center for Advancing Translational Sciences under award number UL1TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire.
[PDF File (Adobe PDF File), 181 KB - jmir_v22i12e18725_app1.pdf]

Multimedia Appendix 2
Interview protocol.
[PDF File (Adobe PDF File), 91 KB - jmir_v22i12e18725_app2.pdf]

Multimedia Appendix 3
Patient perceptions of interpreting test results and the association to health literacy and normality of lab test and patient perceptions of using patient portal and the association to health literacy and technology proficiency.
[DOCX File , 38 KB - jmir_v22i12e18725_app3.docx]

References

1. Lippi G, Mattiuzzi C. Critical laboratory values communication: summary recommendations from available guidelines. Ann Transl Med 2016 Oct;4(20):400 [FREE Full text] [doi: 10.21037/atm.2016.09.36] [Medline: 27867952]
2. Pillemer F, Price R, Paone S, Martich G, Albert S, Haidari L, et al. Direct release of test results to patients increases patient engagement and utilization of care. PLoS One 2016;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
3. Baldwin D, Quintela J, Duclos C, Staton EW, Pace WD. Patient preferences for notification of normal laboratory test results: a report from the ASIPS Collaborative. BMC Fam Pract 2005 Mar 8;6(1):11 [FREE Full text] [doi: 10.1186/1471-2296-6-11] [Medline: 15755328]
4. Ahern D, Woods S, Lightowler M, Finley ST, Houston TK. Promise of and potential for patient-facing technologies to enable meaningful use. Am J Prev Med 2011 May;40(5 Suppl 2):S162-S172. [doi: 10.1016/j.amepre.2011.01.005] [Medline: 21521591]
5. Opennotes and AMIA Join Forces to Improve Patient Access to Health Records. MedCity News. URL: https://medcitynews.com/2017/11/opennotes-amia/ [accessed 2020-11-28]
6. Berlin L. Communicating results of all radiologic examinations directly to patients: has the time come? Am J Roentgenol 2007 Dec;189(6):1275-1282. [doi: 10.2214/ajr.07.2740]
7. Peacock S, Reddy A, Leveille SG, Walker J, Payne TH, Oster NV, et al. Patient portals and personal health information online: perception, access, and use by US adults. J Am Med Inform Assoc 2017 Apr 1;24(e1):e173-e177. [doi: 10.1093/jamia/ocw095] [Medline: 27413120]

8. O'Kane MJ, Lopez B. Explaining laboratory test results to patients: what the clinician needs to know. Br Med J 2015 Dec 3;351:h5552. [doi: 10.1136/bmj.h5552] [Medline: 26634382]
9. Johnson A, Easterling D, Nelson R, Chen MR, Frankel RM. Access to radiologic reports via a patient portal: clinical simulations to investigate patient preferences. J Am Coll Radiol 2012 Apr;9(4):256-263. [doi: 10.1016/j.jacr.2011.12.023] [Medline: 22469376]
10. Walker J, Meltsner MT, Delbanco T. US experience with doctors and patients sharing clinical notes. Br Med J 2015 Mar 10;350:g7785. [doi: 10.1136/bmj.g7785] [Medline: 25670182]
11. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin CT. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
12. O'Kane M, Freedman DB, Zikmund-Fisher BJ. Can patients use test results effectively if they have direct access? Br Med J 2015 Mar 11;350:h673. [doi: 10.1136/bmj.h673] [Medline: 25673132]
13. Witteman HO, Zikmund-Fisher BJ. Communicating laboratory results to patients and families. Clin Chem Lab Med 2019 Feb 25;57(3):359-364 [FREE Full text] [doi: 10.1515/cclm-2018-0634] [Medline: 30407910]
14. Hickner J, Thompson P, Wilkinson T, Epner P, Sheehan M, Pollock AM, et al. Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results. J Am Board Fam Med 2014;27(2):268-274 [FREE Full text] [doi: 10.3122/jabfm.2014.02.130104] [Medline: 24610189]
15. Zikmund-Fisher BJ, Exe NL, Witteman HO. Numeracy and literacy independently predict patients' ability to identify out-of-range test results. J Med Internet Res 2014 Aug 8;16(8):e187 [FREE Full text] [doi: 10.2196/jmir.3241] [Medline: 25135688]
16. McCray A. Promoting health literacy. J Am Med Inform Assoc 2004 Nov 23;12(2):152-163 [FREE Full text] [doi: 10.1197/jamia.m1687]
17. Alpert JM, Krist AH, Aycock RA, Kreps GL. Applying multiple methods to comprehensively evaluate a patient portal's effectiveness to convey information to patients. J Med Internet Res 2016 May 17;18(5):e112 [FREE Full text] [doi: 10.2196/jmir.5451] [Medline: 27188953]
18. Hong M, Feustel C, Agnihotri M, Silverman M, Simoneaux SL, Wilcox L. Supporting Families in Reviewing and Communicating about Radiology Imaging Studies. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2017 May Presented at: CHI'17; May 12-14, 2017; Denver, CO URL: http://europepmc.org/abstract/MED/28920107 [doi: 10.1145/3025453.3025754]
19. Sarkar U, Karter A, Liu J, Adler N, Nguyen R, Lopez AD, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15(Suppl 2):183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
20. Tang P, Overhage J, Chan A, Brown N, Aghighi B, Entwistle M, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 1;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
21. Zikmund-Fisher BJ, Scherer AM, Witteman HO, Solomon JB, Exe NL, Fagerlin A. Effect of harm anchors in visual displays of test results on patient perceptions of urgency about near-normal values: experimental study. J Med Internet Res 2018 Mar 26;20(3):e98 [FREE Full text] [doi: 10.2196/jmir.8889] [Medline: 29581088]
22. Peters E, Dieckmann N, Dixon A, Hibbard JC, Mertz CK. Less is more in presenting quality information to consumers. Med Care Res Rev 2007 Apr;64(2):169-190. [doi: 10.1177/10775587070640020301] [Medline: 17406019]
23. Giardina T, Baldwin J, Nystrom D, Sittig DH, Singh H. Patient perceptions of receiving test results via online portals: a mixed-methods study. J Am Med Inform Assoc 2018 Apr 1;25(4):440-446 [FREE Full text] [doi: 10.1093/jamia/ocx140] [Medline: 29240899]
24. Christensen K, Sue V. Viewing laboratory test results online: patients' actions and reactions. J Particip Med 2013;5:e38.
25. Reynolds TL, Ali N, McGregor E, O'Brien T, Longhurst C, Rosenberg AL, et al. Understanding patient questions about their medical records in an online health forum: opportunity for patient portal design. AMIA Annu Symp Proc 2017;2017:1468-1477 [FREE Full text] [Medline: 29854216]
26. Zhang Z, Lu Y, Kou Y, Wu DT, Huh-Yoo J, He Z. Stud Health Technol Inform 2019 Aug 21;264:1403-1407 [FREE Full text] [doi: 10.3233/SHTI190458] [Medline: 31438157]
27. Wilcox L, Patel R, Chen Y, Shachak A. Human factors in computing systems: focus on patient-centered health communication at the ACM SIGCHI conference. Patient Educ Couns 2013 Dec;93(3):532-534. [doi: 10.1016/j.pec.2013.09.017] [Medline: 24184039]
28. Reiff M, Bernhardt BA, Mulchandani S, Soucier D, Cornell D, Pyeritz RE, et al. 'What does it mean?': uncertainties in understanding results of chromosomal microarray testing. Genet Med 2012 Mar;14(2):250-258 [FREE Full text] [doi: 10.1038/gim.2011.52] [Medline: 22241091]
29. Nystrom DT, Singh H, Baldwin J, Sittig DF, Giardina TD. Methods for patient-centered interface design of test result display in online portals. EGEMS (Wash DC) 2018 Jun 26;6(1):15 [FREE Full text] [doi: 10.5334/egems.255] [Medline: 30094287]

30. Denscombe M. Communities of practice. J Mix Method Res 2008 Jul;2(3):270-283 [FREE Full text] [doi: 10.1177/1558689808316807]
31. Mason J. Mixing methods in a qualitatively driven way. Qual Res 2016 Aug 15;6(1):9-25 [FREE Full text] [doi: 10.1177/1468794106058866]
32. Axinn WG, Pearce LD. Mixed Method Data Collection Strategies. Cambridge, England: Cambridge University Press; 2006.
33. Shapiro D, Chandler JP, Mueller PA. Using mechanical Turk to study clinical populations. Clin Psychol Sci 2013 Jan 31;1(2):213-220 [FREE Full text] [doi: 10.1177/2167702612469015]
34. Taha J, Sharit J, Czaja SJ. The impact of numeracy ability and technology skills on older adults' performance of health management tasks using a patient portal. J Appl Gerontol 2014 Jun;33(4):416-436 [FREE Full text] [doi: 10.1177/0733464812447283] [Medline: 24781964]
35. Elder NC, Barney K. 'But what does it mean for me?' Primary care patients' communication preferences for test results notification. Jt Comm J Qual Patient Saf 2012 Apr;38(4):168-176. [doi: 10.1016/s1553-7250(12)38022-7] [Medline: 22533129]
36. Braun VV, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
37. Beyer HK, Holtzblatt K. Contextual design: defining customer-centered systems. New York: Elsevier; 1998:978.
38. Chen Y, Cheng K, Tang C, Siek KA, Bardram JE. Is My Doctor Listening to Me? Impact of Health IT Systems on Patient-provider Interaction. In: Extended Abstracts on Human Factors in Computing Systems. 2013 Presented at: CHI'13; April 27, 2013; Paris, France URL: https://doi.org/10.1145/2468356.2468791 [doi: 10.1145/2468356.2468791]
39. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
40. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
41. Miller DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary care providers' views of patient portals: interview study of perceived benefits and consequences. J Med Internet Res 2016 Jan 15;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
42. Park MS, He Z, Chen Z, Oh S, Bian J. Consumers’ use of UMLS concepts on social media: diabetes-related textual data analysis in blog and social Q&A sites. JMIR Med Inform 2016 Nov 24;4(4):e41 [FREE Full text] [doi: 10.2196/medinform.5748] [Medline: 27884812]
43. Yu B, He Z, Xing A, Lustria ML. An informatics framework to assess consumer health language complexity differences: proof-of-concept study. J Med Internet Res 2020 May 21;22(5):e16795 [FREE Full text] [doi: 10.2196/16795] [Medline: 32436849]
44. Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. SPIKES-A six-step protocol for delivering bad news: application to the patient with cancer. Oncologist 2000;5(4):302-311 [FREE Full text] [doi: 10.1634/theoncologist.5-4-302] [Medline: 10964998]
45. Curioni-Fontecedro A. A new era of oncology through artificial intelligence. ESMO Open 2017;2(2):e000198 [FREE Full text] [doi: 10.1136/esmoopen-2017-000198] [Medline: 28761754]
46. Korevaar T. The upper limit for TSH during pregnancy: why we should stop using fixed limits of 2.5 or 3.0 mU/l. Thyroid Res 2018;11:5 [FREE Full text] [doi: 10.1186/s13044-018-0048-7] [Medline: 29942352]
47. Choe EK, Duarte ME, Suh H, Pratt W, Kientz JA. Communicating bad news: insights for the design of consumer health technologies. JMIR Hum Factors 2019 May 17;6(2):e8885 [FREE Full text] [doi: 10.2196/humanfactors.8885] [Medline: 31102374]
48. Ferrante J, Chen PS, Kim S. The effect of patient navigation on time to diagnosis, anxiety, and satisfaction in urban minority women with abnormal mammograms: a randomized controlled trial. J Urban Health 2008 Jan;85(1):114-124 [FREE Full text] [doi: 10.1007/s11524-007-9228-9] [Medline: 17906931]
49. Centers for Disease ControlPrevention (CDC), HHS, Office for Civil Rights (OCR), HHS. CLIA program and HIPAA privacy rule; patients' access to test reports. Final rule. Fed Regist 2014 Mar 6;79(25):7289-7316 [FREE Full text] [Medline: 24605389]
50. Giardina TD, Modi V, Parrish DE, Singh H. The patient portal and abnormal test results: an exploratory study of patient experiences. Patient Exp J 2015;2(1):148-154 [FREE Full text] [Medline: 28345018]
51. Kelman A, Robinson CO, Cochin E, Ahluwalia NJ, Braverman J, Chiauzzi E, et al. Communicating laboratory test results for rheumatoid factor: what do patients and physicians want? Patient Prefer Adherence 2016;10:2501-2517 [FREE Full text] [doi: 10.2147/PPA.S104396] [Medline: 28008236]
52. Martinez W, Threatt AL, Rosenbloom ST, Wallston KA, Hickson GB, Elasy TA. A patient-facing diabetes dashboard embedded in a patient web portal: design sprint and usability testing. JMIR Hum Factors 2018 Sep 24;5(3):e26 [FREE Full text] [doi: 10.2196/humanfactors.9569] [Medline: 30249579]

Abbreviations

AI: artificial intelligence
CMA: chromosomal microarray
MTurk: Amazon Mechanical Turk
NIH: National Institutes of Health
RQ: research question

Edited by G Eysenbach; submitted 14.03.20; peer-reviewed by D He, J Zekri; comments to author 12.06.20; revised version received 11.08.20; accepted 11.11.20; published 07.12.20.

Please cite as:
Zhang Z, Citardi D, Xing A, Luo X, Lu Y, He Z
Patient Challenges and Needs in Comprehending Laboratory Test Results: Mixed Methods Study
J Med Internet Res 2020;22(12):e18725
URL: http://www.jmir.org/2020/12/e18725/
doi:10.2196/18725
PMID:33284117

Review

Health Care Professionals' Experiences of Patient-Professional Communication Over Patient Portals: Systematic Review of Qualitative Studies

Elina Laukka[1,2], RN, MSc; Moona Huhtakangas[2], MSc; Tarja Heponiemi[1], PhD; Sari Kujala[3], PhD; Anu-Marja Kaihlanen[1], RN, PhD; Kia Gluschkoff[1,4], PhD; Outi Kanste[2], RN, PhD

[1]Finnish Institute for Health and Welfare, Social and Health System Research Unit, Helsinki, Finland
[2]Research Unit of Nursing Science and Health Management, University of Oulu, Oulu, Finland
[3]Department of Computer Science, Aalto University, Espoo, Finland
[4]Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

Corresponding Author:
Elina Laukka, RN, MSc
Finnish Institute for Health and Welfare
Social and Health System Research Unit
Mannerheimintie 166
Helsinki, 00271
Finland
Phone: 358 29 524 7368
Email: elina.laukka@thl.fi

Abstract

Background: The popularity of web-based patient-professional communication over patient portals is constantly increasing. Good patient-professional communication is a prerequisite for high-quality care and patient centeredness. Understanding health care professionals' experiences of web-based patient-professional communication is important as they play a key role in engaging patients to use portals. More information is needed on how patient-professional communication could be supported by patient portals in health care.

Objective: This systematic review of qualitative studies aims to identify how health care professionals experience web-based patient-professional communication over the patient portals.

Methods: Abstract and full-text reviews were conducted by 2 reviewers independently. A total of 4 databases were used for the study: CINAHL (EBSCO), ProQuest (ABI/INFORM), Scopus, and PubMed. The inclusion criteria for the reviewed studies were as follows: the examination of health care professionals' experiences, reciprocal communication between patients and health care professionals, peer-reviewed scientific articles, and studies published between 2010 and 2019. The Joanna Briggs Institute's quality assessment criteria were used in the review process. A total of 13 included studies were analyzed using a thematic synthesis, which was conducted by 3 reviewers.

Results: A total of 6 analytical themes concerning health care professionals' experiences of web-based patient-professional communication were identified. The themes were related to health care professionals' work, change in communication over patient portals, patients' use of patient portals, the suitability of patient portals for communication, the convenience of patient portals for communication, and change in roles.

Conclusions: Health care professionals' experiences contain both positive and negative insights into web-based patient-professional communication over patient portals. Most commonly, the positive experiences seem to be related to the patients and patient outcomes, such as having better patient engagement. Health care professionals also have negative experiences, for example, web-based patient-professional communication sometimes has deficiencies and has a negative impact on their workload. These negative experiences may be explained by the poor functionality of the patient portals and insufficient training and resources. To reduce health care professionals' negative experiences of web-based patient-professional communication, their experiences should be taken into account by policy makers, health care organizations, and information technology enterprises when developing patient portals. In addition, more training regarding web-based patient-professional communication and patient portals should be provided to health care professionals.

(J Med Internet Res 2020;22(12):e21623) doi:10.2196/21623

KEYWORDS

telemedicine; communication; patient portals; nurses; physicians; systematic literature review; thematic analysis

Introduction

Background

Due to the World Health Organization's aims for health care digitalization and due to the COVID-19 pandemic, the importance of eHealth has increased considerably [1,2]. Internet-based interactive health services such as patient portals represent one form of eHealth [3]. Patient portals are secure websites that offer patients access to a variety of functions, such as viewing laboratory results and secure messaging; these portals are administrated and owned by health care organizations [4,5]. Patient portals provide patients with remote web-based access to their personal health information, services, and clinical care [3], and occasionally patient portals are synchronized with electronic health records (EHRs) [4]; however, they may also be individual web pages with no connection to EHRs. Occasionally, portals enable reciprocal communication between patients and health care professionals [6,7], for example, via secure electronic messaging [8].

Good patient-professional communication is a prerequisite for high-quality care and a key element of patient centeredness [9]. Traditionally, patient-professional communication happened face-to-face during clinical consultations [8]; however, in recent years, communication through patient portals has also become common [10]. Patient portals enable reciprocal communication and interactive guidance and coaching of patients, which may be more effective than just providing patients with clinical information, such as doctors' notes, without any further advice [4]. Some patients even prefer to use web-based communication, for example, because it can be seen as less intimidating than face-to-face encounters [11,12]. Patients have reported several benefits related to the use of patient portals. For example, patients with diabetes were significantly more likely to believe that reading their doctor's notes would improve their self-care and adherence to medication [13]. In addition, web-based communication with patients with cancer may have just as big of an impact on their care as face-to-face communication [14]. The portals not only helped patients to better manage their diseases but also conferred psychological benefits, such as increasing trust and collaboration with health care providers [15].

Patient portals provide several benefits for service providers, such as reduced amounts of hospital-based care [16] and cost-effectiveness [17]. However, many physicians have expressed concerns that the use of patient portals could change the patient-professional relationship [18]. In a study by Daniel et al [19], the majority of physicians showed reluctance to use web-based apps and social media to communicate with patients. In addition, some health care professionals have been unwilling to inform patients about patient portals and have expressed concern that patient portals may reduce their professional autonomy [20]. Despite concerns and reluctance of health care professionals, digitalization and web-based communication have already become a part of health care professionals' expected competency [21].

Traditional face-to-face patient-professional communication has been widely examined [9,22,23], but less attention has been paid to web-based patient-professional communication that occurs over patient portals [24]. Previous studies have shown conflicting results related to web-based patient-professional communication. In a systematic review, Kruse et al [25] found that several positive and negative attributes of web-based patient-professional communication overlapped within the same study. For example, although some patients and professionals perceived an element of a patient portal to be beneficial, other respondents had negative experiences related to the same element in their portal [25]. Another systematic review supports the nature of bifurcation; Ferreira et al [26] found that some studies enhanced patient-professional communication but also showed patients' concerns about confidentiality and understanding of the content.

Web-based patient-professional communication has been synthesized in a scoping review by Voruganti et al [27], who aimed to map, describe, and understand web-based tools for communication between patients and physicians. They found that web-based tools for patient-professional communication were most prevalent in contexts where the intended use was the patients' self-management [27]. In this review, the experiences of health care professionals will be scrutinized to gain broader knowledge about web-based patient-professional communication. This review focuses on patient portals, as they seem to be the most commonly used web-based tools for patient-professional communication [27].

Understanding the experiences of health care professionals with patient portals is important because health care professionals play a key role when supporting and engaging patients in their use of these portals [20]. Endorsement of patient portals by health care professionals is one of the most influential factors impacting patients' use of patient portals as well as their use as tools for collaborative communication [28]. Aligning patient portals with health care professionals' workflow and care delivery priorities is difficult, and this might impact the professionals' experiences of the patient-professional communication over patient portals [29]. According to Irizarry et al [29], greater understanding is needed of how patient-professional communication could be supported by patient portals in practical care work.

Objectives

As previous reviews have found conflicting results [25] and concentrated on web-based tools instead of communication [27], a more detailed understanding is needed on patient-professional communication from a professional perspective. Adding this information may support the use of patient portals in practical care work. In light of these gaps in the research, a systematic literature review was conducted to identify the experiences of health care professionals with web-based patient-professional communication over patient portals. In this study, we define communication as reciprocal web-based communication or interaction between health care professionals and patients. With reciprocal, we mean that patients can answer their health care professional directly using, for example, secure messaging. The following research question was addressed: What kind of experiences do health care professionals have of web-based patient-professional communication over patient portals?

Methods

Information Sources and Search Strategy

A systematic review of qualitative studies was conducted following the Joanna Briggs Institute's (JBI) Reviewer's Manual [30], including the application of a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [31]. A systematic review method was chosen for this review because it is applicable in areas where there is only a little preexisting knowledge and where complex issues require more detailed exploration [32]. With the assistance of an experienced information specialist, the searches were conducted in the CINAHL (EBSCO), ProQuest (ABI/INFORM), Scopus, and PubMed databases using search terms related to portals and patient-professional communication (Table 1). A total of 1038 articles were found, which were reduced to 597 after duplicates were removed using the RefWorks Legacy reference management software package.

Table 1. Databases, search strategy, and results identifying the studies.

Database	Search strategy	Results, n
Cumulative index to nursing and allied health literature (CINAHL [EBSCO])	ALL[a](portal OR portals) AND ALL(patient* N5 (professional* OR provider* OR physician* OR doctor* OR nurse*)) AND ALL(communicat* OR interact*)	183
ProQuest (ABI/INFORM)	ALL(portal OR portals) AND ALL(patient* N/5 (professional* OR provider* OR physician* OR doctor* OR nurse*)) AND ALL(communicat* OR interact*)	12
Scopus	ALL(portal OR portals) AND ALL(patient* W/5 (professional* OR provider* OR physician* OR doctor* OR nurse*)) AND ALL(communicat* OR interact*)	382
PubMed	TEXT WORD[b](portal OR portals) AND TEXT WORD(communicat* OR interact*) AND TEXT WORD(patient* AND (professional OR provider OR physician OR doctor* OR nurse*)	461

[a] ALL: Everywhere but the whole text, that is, including the title, abstracts, and keywords.

[b] TEXT WORD: Terms that are qualified with the Text Word field tag will be searched in the following fields: title, abstract, medical subject headings (MeSH) heading and subheadings, other terms field, secondary source identifier, and personal name as subject.

Study Selection Process

The study selection process is presented in the PRISMA flow diagram shown in Figure 1. The studies were screened by title and abstract (n=597) and the full text (n=53) independently by 2 researchers (EL and MH). The inclusion was based on eligibility criteria that were defined according to the participants, phenomenon of Interest, context, study type method: (1) the *P*articipants were health care professionals; (2) the phenomenon of *I*nterest was communication; (3) the *C*ontext referred to patient portals; and (4) as a *S*tudy type, only studies including qualitative data were included because they better examined experiences (Table 2). All the studies were published between 2010 and 2019. This time span is relevant for researching patient portals, which are a relatively new phenomenon and are constantly developing. The 2 reviewers discussed and agreed on which studies should be included according to the inclusion criteria. A total of 41 articles were excluded because they did not meet the eligibility criteria. Moreover, 12 of the excluded articles were from the patient, manager, or caregiver perspective. In 19 articles, the phenomenon of interest was not in patient-professional communication, but was focused, for example, on scrutinizing the communication between professionals or the communication was not reciprocal. In addition, 3 articles scrutinized web-based patient-professional communication in emails and open notes instead of patient portals. Finally, 8 studies represented the wrong study type or the full text was not available. The reasons for exclusions are presented in Figure 1. The reference lists of all the included studies (N=13) were manually searched for additional studies (n=1). If any disagreements occurred at any point in the article selection process, they were resolved by consulting the last author of this paper (OK). A kappa value of 0.72 in the title-abstract screening and 0.95 in full-text screening showed a substantial and almost perfect level of agreement [33].

Figure 1. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the systematic review carried out in this study.

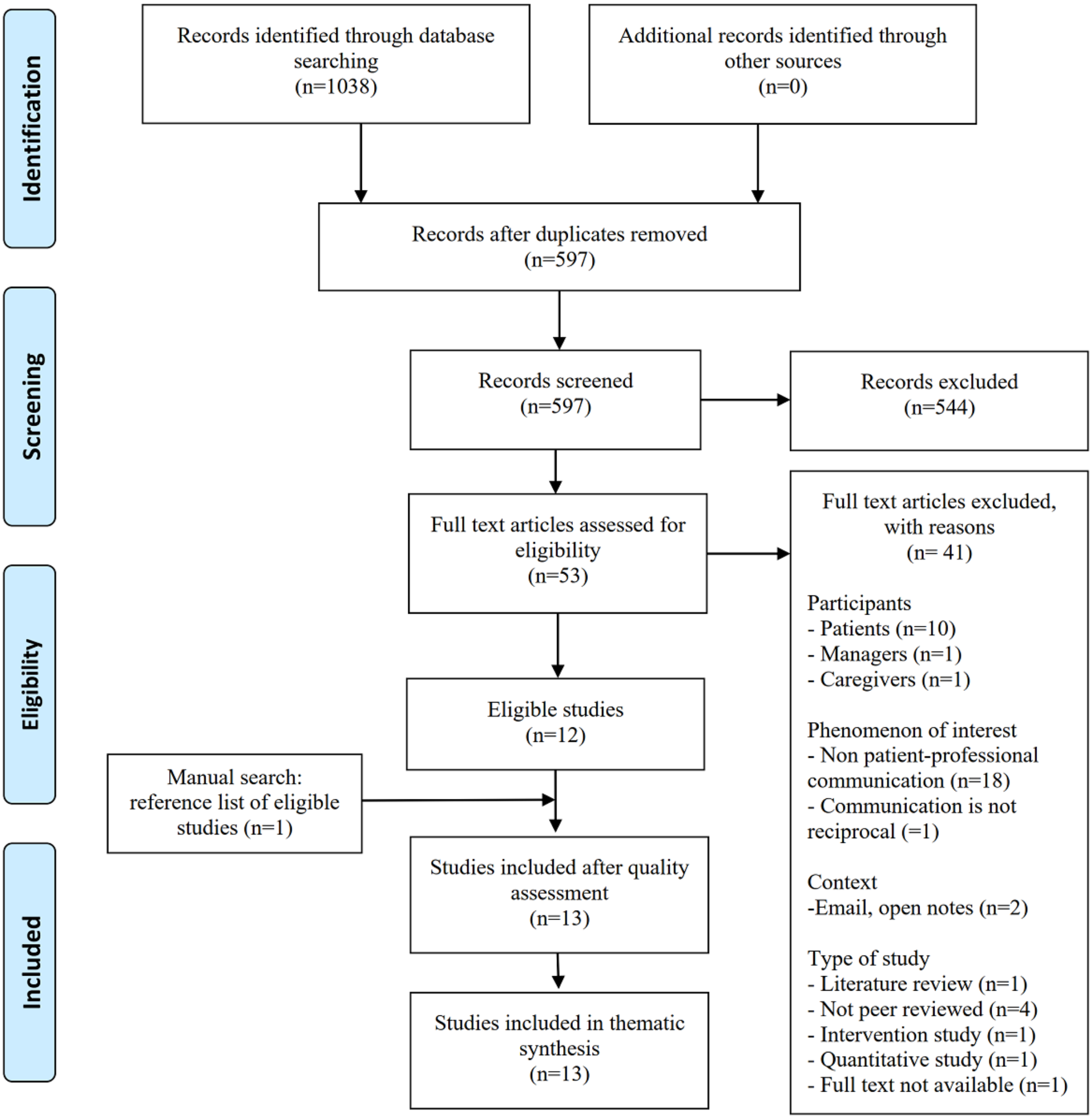


Table 2. Eligibility criteria according to the PICoS protocol.

PICoS[a]	Inclusion criteria	Exclusion criteria
Participants	Health care professionals	Other than health care professionals (eg, managers or patients)
Phenomenon of interest	Reciprocal web-based communication or interaction between health care professionals and patients	Communication occurs between health care professionals and communication occurs between health care professional and family member or caregiver
Context	Patient portal	Other than patient portal (eg, email, video consultations, and open notes)
Types and quality of studies	Peer-reviewed scientific studies containing qualitative data, published between 2010 and 2019, full text available	Quantitative studies, gray literature, literature reviews, systematic reviews, study protocols, and intervention studies

[a] PICoS: participants, phenomenon of interest, context, study.

Quality Assessment

The quality of the studies (N=13) chosen for the review was assessed by 2 independent reviewers (EL and MH) using the JBI checklist for qualitative research [34]. To achieve consensus between the reviewers, the selected studies had to achieve at least a score of 5 out of 10 across the quality criteria [35]. All 13 studies were considered suitable for the review.

Data Extraction and Synthesis

The data were extracted by the authors, year of publication, country of origin, purpose, participants, methods (data collection and analysis), relevant findings, and quality assessment (Table 3). The thematic synthesis presented by Thomas and Harden [36] was used to synthesize the results. The synthesis included line-by-line coding of the findings, organizing initial codes (n=162) to construct descriptive themes (n=48), the use of subthemes when necessary to clarify the reporting of the results (n=11), and finally the development of analytical themes (n=6) [36,37]. An example of coding and theme building is presented in Table 4.

Table 3. Data extraction of the included studies.

Reference, country of origin	Purpose	Participants (relevant participants for the review)	Methods (data collection and analysis)	Relevant findings	Quality assessment points
Alpert et al [38], United States	To evaluate how well portals convey information to patients. To demonstrate how methodologies could be used to evaluate and improve the design of portals.	Clinicians (n=13), patients (n=31).	Interviews of patients, focus groups for clinicians (n=2). Thematic analysis.	Clinicians believed that the portal was not well equipped to handle complex communication. They found it complicated that there was no confirmation that the patients viewed the messages. Asynchronous communication patterns disrupted care.	7/10
Alpert et al [39], United States	To describe the types of messages initiated by patients communicating via patient portals and to assess whether providers employed patient-centered strategies in their responses.	N/A[a]	A total of 193 messages from 58 message threads. Content analysis.	Providers limited their responses strictly to the requested information, and the majority of their responses lacked empathy. Occasionally, providers adopted a "customer service-oriented" approach. Some of the providers' messages reinforced positive patient behavior.	6/10
Alpert et al [40], United States	To understand attitudes of the portal's adoption for oncology and to identify the advantages and disadvantages of using the portal to communicate and view medical information.	Oncologists (n=13), medical informaticists (n=12), and patients (n=35).	In-depth semistructured interviews. Thematic text analysis.	Oncologists believed that advanced access may improve engagement during consultations and also assist patients after appointments.	7/10
Alpert et al [41], United States	To understand the perceptions of oncologists and cancer patients about the potential impact of portals on such communication.	13 oncologists and 35 patients.	In-depth, semistructured interviews. Thematic analysis.	Oncologists agreed that patient access to their records was beneficial, but they were also concerned about the workload and portal not being suitable for complex information. Oncologists were also concerned that patients would anticipate rapid communication and that they would learn about a new diagnosis before meeting the oncologist.	8/10
Bishop et al [42], United States	To answer the following research questions: (1) how can primary care practices use electronic communication to manage clinical issues; (2) what are the perceived advantages and disadvantages of these programs for patients, physicians, and practices; and (3) what are the barriers to and facilitators of implementation of electronic communication programs?	Leaders of 21 medical groups and also staff in 6 of these groups.	Interviews analyzed using a constant comparative method.	Frontline physicians agreed that electronic communication improved access to care for patients, saved patients' time, and improved patient satisfaction. They also cited efficiency as an advantage of patient portals. Physicians were able to contact patients before the consultations, which improved the efficiency of office visits. Increased workload and patient and physician resistance were identified as disadvantages.	6/10
Das et al [43], Norway	To characterize and assess the impact of an eHealth portal on health care professionals' interaction with patients in bariatric surgery.	Health care professionals (n=3).	Semistructured in-depth interviews that were thematically analyzed.	By following the patients' writing, the professionals learned more about their patients than during time-limited face-to-face consultations. The portal became an important way to reach out for the patients. Yet, professionals reported uncertainty about how to deal with new kinds of interaction, and they were not able to ensure that patients understood the given information. The portals also increased the workload, interfered with the workflow, and were not suitable for complex cases.	7/10

Reference, country of origin	Purpose	Participants (relevant participants for the review)	Methods (data collection and analysis)	Relevant findings	Quality assessment points
Elers and Nelson [44], New Zealand	To identify how general practitioners perceive patient portals to influence the delivery of primary health care.	General practitioners (n=9).	Semistructured interviews that were thematically analyzed.	General practitioners expressed that the internet was just another way to interact with the patients.	7/10
Gerber et al [45], United States	To identify nursing staff reactions to and perceptions of electronic portal use in a cancer setting.	Outpatient nurses (n=13).	Focus groups (n=2). Theoretical thematic content analysis.	Nurses were concerned about the increase in the volume of electronic communications and that patients expected immediate responses.	8/10
Kopanitsa [46], Russia	To analyze the attitudes of patients with tuberculosis and doctors and identify perceived opportunities and barriers to operate a web portal.	General practitioners (n=10), phthisiatricians (n=8), and patients (n=30).	Semistructured interviews. Grounded theory and thematic analysis.	Doctors reported that the web portal would improve their communication with patients, but should not create any additional work.	8/10
Miller et al [47], United States	To determine how administrators, clinic staff, and health care providers at a practice serving a lower-income adult population viewed patient portals in terms of their potential benefit, areas of concerns, and hopes for the future.	Clinical personnel (n=20).	In-depth interviews. Data analysis was based on a systematic, computer-assisted approach.	Nurses, physicians, and clinic personnel agreed that electronic messaging was quicker and more efficient. Several nurses and physicians were worried that some patients would inappropriately send repeated messages and may expect immediate responses to their electronic requests. Health care professionals were aware that patients saw all the posts.	7/10
Nazi [48], United States	To examine the experiences of physicians, nurses, and pharmacists at the Department of Veterans Affairs using an organizationally sponsored personal health records to develop insights into the interaction of technology and processes of health care delivery.	Health care professionals (n=30).	In-depth interviews. Utilized modern techniques of qualitative analysis.	Secure messaging enabled better connectivity between patients and their health care teams. Asynchrony was seen as a benefit that allowed health care professionals to send and respond to messages when it was convenient for them. Secure messaging enabled health care professionals to better know their patients and enhanced the quality of visits.	7/10
Sieck et al [49], United States	Within primary care offices with high rates of patient portal use, the study examined how experienced physicians and patient users of the ambulatory portal perceived the benefits and challenges of portal use in general and secure messaging in particular.	Primary care physicians (n=13) and patients (n=29).	Semistructured interviews. Both inductive and deductive methods, using a constant comparative analytic approach.	Providers noted increased efficiency in communications, but some of the problems were too complex to handle via secure messaging. Sometimes the providers noted that patient messages did not contain enough information, and hence, they were worried about the quality of the provided information.	8/10
Vreugdenhil et al [50], Netherlands	To explore the adoption, use, usability, and usefulness of a recently introduced patient portal in an academic hospital to learn lessons for the implementation of patient portals in fragmented health care systems.	Medical specialists (n=3), medical specialists in training (n=4), nurses (n=4), administrative assistants (n=3), doctor’s assistant (n=1), managers (n=2), and patients (focus group n=7, think-aloud observation n=8).	Focus groups (n = 4) and think-aloud observations for patients. Thematic content analysis.	Not all the health care professionals agreed on how the messaging functionality should be used. Some doctors preferred not to use messaging functionality to answer questions, especially when it concerned complex problems. Health care professionals felt they lost some control because of the portal.	7/10

[a] N/A: not applicable.

Table 4. An example of coding in thematic synthesis.

Line-by-line coding[a]	Initial code[b]	Descriptive theme[c]	Subtheme[d]	Analytical theme[e]
"It is just that the days are filled with patient lists, and suddenly it is 4 o'clock, and then you are off to home. We haven't organized the time for it..." [43]	No time organized for portal communication	Not enough time resources	Lack of time and expertise for portal communication	Health care professionals' work
"The way my day is set up right now, I am scheduled to see patients; I really have no time to respond..." [42]	No time to respond	__[f]	—	—
"You need to plan extra time to process these messages." [50]	Need for extra time	—	—	—
"'Linda' explained that the activities triggered by this one question required considerable effort: the process required resources in regards to expertise in knowing the right addressee, time and effort to contact them..." [43]	Time effort	—	—	—
"...time constraints, and prioritizations became evident in the daily clinical practice..." [43]	Time constraints	—	—	—
"This represented a challenge for the level of expertise required..." [43]	Challenges due to the level of expertise required	Lack of expertise	—	—
"In cases when the personnel with portal access could not respond themselves, they made contact with other professionals at the clinic..." [43]	Professionals could not answer themselves	—	—	—

[a] Free line-by-line coding of the findings of primary studies.

[b] Initial codes based on line-by-line coding and formed into a *bank* of codes.

[c] Initial codes categorized as descriptive themes.

[d] Descriptive themes were grouped under a subtheme.

[e] Sufficiently abstract analytical themes were created to describe all the descriptive themes. This table does not contain all the descriptive themes that were categorized under Health care professionals' work.

[f] The empty cells are not supposed to contain text.

First, free line-by-line coding was performed for the findings of the primary studies. Second, initial codes were created using line-by-line coding. Every sentence had at least one initial code, but using several codes for a sentence was also possible, albeit rare, in this review. The use of line-by-line coding enabled the translation of concepts from one study to another. Initial codes formed a *bank* of codes, and new codes were developed when necessary, and the synthesis began at this point. Third, codes were grouped into descriptive themes based on their similarity. Heretofore, the synthesis was kept very close to the original findings of the included studies. Fourth, some of the descriptive themes were grouped under subthemes. Finally, analytical themes were formed. After this phase, the analytical themes were sufficiently abstract to describe all the initial descriptive themes (Table 5) [36].

Table 5. An overview of analytical themes, subthemes, and descriptive themes (number of studies out of the included 13).

Analytical themes (n=6)	Subthemes (11)	Descriptive themes (n=48)
Health care professionals' work (13)	Increased workload	• Increases workload (8) • Causes more steps in the care process (6) • Causes additional tasks (1)
	More efficient work	• Increases efficiency of work (6) • Saves time (2) • Reduces telephone communication (2)
	Experiences of fear and discomfort	• Causes fear (3) • Confusion in professional activities (2) • Uncertainty about portal communication (1) • Increases pressure to contact patients faster who actively use the portal (1)
	Increased awareness of the patients' situations	• Provides more information about the patients (3) • Enables better overall impression of the patients' situations (2)
	Lack of time and expertise for engaging in portal communication	• Not enough time or resources (3)
Change in communication over the patient portals (12)	Enhanced communication	• Enables more direct communication (6) • Enhances connectivity (2) • Promotes more frequent communication (2) • More focused attention (2) • Easy way to communicate (2) • Safe means of communication (2) • Improves communication and relationship (1)
	Change in the means of communication	• Transforms the means of communication (5) • Variable nature of communication (1)
	Deficiencies in communication	• Inability to communicate appropriately (4) • Insensitive communication (2) • General nature of information, not detailed (2)
Patients' use of patient portals (12)	Interpretation and communication	• Uncertainty about whether patients understand the information (8) • Patients communicating inappropriately (3) • Patients providing insufficient information (1)
	Positive consequences for patients	• Enables patient outcomes and experiences (6) • Promotes patients' involvement (4) • Lowers threshold of communication (2) • Nonstigmatic way to communicate (1) • Patients ask more specific questions (1)
	Patients' high expectations of professionals	• Patients' expectations for rapid communication (3) • Patients' expectations for use of data (1)
Suitability of the patient portals for communication (7)	N/A[a]	• Not suitable for complex communication (3) • Not suitable for complex cases (3) • Useful in remotely managing patients' conditions (2) • Not suitable for complex communication (1) • Uniqueness of patient cases (1)
Convenience of the patient portals for communication (7)	N/A	• Flexibility to answer when convenient (4) • Possibility to asynchronous communication (3) • Portal seen as a nuisance (3) • Uncertainty if patients receive the information (1)

Analytical themes (n=6)	Subthemes (11)	Descriptive themes (n=48)
Change in roles (4)	N/A	• Changes patients' roles (2) • Maintains health care professionals' responsibility for the patients (2) • Transforms health care professionals' roles (1)

[a]N/A: not applicable.

The synthesis was conducted independently by 3 reviewers (EL, MH, and OK) at all stages. After each stage, the reviewers looked for similarities and differences in their codes and themes and agreed on the final versions.

Results

Study Characteristics

The included studies (N=13) originated from the United States (n=9, 69%), Norway (n=1, 8%), the Netherlands (n=1, 8%), New Zealand (n=1, 8%), and Russia (n=1, 8%). The informants in the included studies were reported to be physicians (n=82), health care professionals or clinical personnel (n=63), nurses (n=17), and medical specialists in training (n=4). The number of informants was not reported in 2 studies, but it was reported that they were health care professionals. The data were most commonly collected using individual interviews (n=10), but focus groups (n=2), message threads (n=1), and observation (n=1) were also used for collecting data.

Health Care Professionals' Experiences of Web-Based Patient-Professional Communication Over Patient Portals

When analyzing the included studies (N=13), 6 analytical themes were identified that were experiences related to (1) the health care professionals' work (13/13, 100%), (2) changes in communication over the patient portals (12/13, 92%), (3) the patients' use of patient portals (12/13, 92%), (4) the suitability of patient portals for communication (7/13, 54%), (5) the convenience of the patient portals for communication (7/13, 54%), and (6) the change in roles (4/13, 8%; Table 5).

Experiences Related to Health Care Professionals' Work

Experiences related to health care professionals' work were divided into 5 subthemes, which related to the increased workload [39,41,42,44-48,50], more efficient work [42,44,47,48,50,51], experiences of fear and discomfort [41,43,46,49,50], increased awareness of patients' situations [41,43,48], and lack of time and expertise required for portal communication [42,43,50].

Increased Workload

The health care professionals in the studies were concerned about the patient-professional communication increasing their workload [41,42,44,46,48,50]. Although they saw patient participation as a positive outcome of portal communication, having greater patient participation increased the workload [41]. Sometimes health care professionals received messages from patients who did not actually need their help [46], and some of the patients overwhelmed the health care professionals by inappropriately sending them repeated messages [47]. There were organizations where answering the portal messages was not formally incorporated into the daily work process of the professionals, and thus, some of the health care professionals had multiple portal messages waiting for responses, causing them an additional workload [43,50].

Web-based patient-professional communication also increased the steps in the care process [38-40,43,45,47]. For example, web-based communication sometimes triggered more phone calls and several follow-up questions [38,47]. Thus, health care professionals occasionally found it easier to just recommend that the patients scheduled an appointment instead of communicating on the internet [39]. Health care professionals could not always trust that the patient had truly received the information over the patient portal [38]. Thus, the professionals had to ensure by phone that patients really had received the information [38]. When answering the patients, health care professionals also felt a need to ensure that the content of their message was correct and that the patient would not misunderstand the core message [43]. This led to another extra step in the process.

More Efficient Work

Some health care professionals reported that web-based patient-professional communication increased the efficiency of their work [42,44,47,48,50,51]. Not only was the web-based communication itself more effective but it also improved the efficiency and quality of face-to-face office visits because patients were able to communicate with professionals before the visits [42,47,48,50]. Face-to-face visits were improved because communicating with the patients on the internet before the visits helped the professionals to be more prepared for the consultation [42]. According to some health care professionals, communicating over the portal also saves time because sending electronic messages is faster than making phone calls [47]. Due to the web-based patient-professional communication, health care professionals were able to reduce the volume of incoming phone calls [47], thus avoiding some challenges encountered with them [48].

Experiences of Fear and Discomfort

In some cases, communicating over the patient portals caused a degree of fear and concern among health care professionals. Health care professionals feared that patients may discover a significant change in their well-being, diagnosis, or prognosis in the doctors' notes without first communicating with health care professionals [41,46]. Due to this fear, the professionals felt pressure to contact patients who actively used portals faster to prevent them from discovering any changes by themselves [41].

In some cases, health care professionals were afraid of communicating on the internet [43]. Patient portals caused some confusion in the professionals' activities [49,50] by making health care professionals feel slightly uncomfortable [50] or *getting lost* in their professional activities [49]. In addition, health care professionals felt uncertain about how to deal with portal communication, and they felt that they had to focus on the articulation and content of their messaging, which again required considerable effort [43].

Increased Awareness of Patients' Situations

Web-based patient-professional communication increased the health care professionals' knowledge of their patients [41] and resulted in receiving more detailed patient information [43]. On patient portals, health care professionals could better capture things that did not come up during face-to-face consultations [43], and patients were more inclined to share sensitive information [48]. Due to web-based patient-professional communication, health care professionals gained a better overall impression of their patients' situations [43,48]. Due to the more frequent communication, health care professionals knew their patients better [48] and learned more about their patients' needs [43].

Lack of Time and Expertise for Engaging in Portal Communication

Sometimes, health care professionals reported that they did not have enough time for portal communication [42,43,50], and they needed extra time to process the portal messages [50]. For example, some health care professionals answered messages in their spare time [50]. The professionals were also challenged by the level of expertise required for portal communication, as they were not always able to answer portal messages by themselves and needed support from other professionals [43].

Experiences Related to Changes in Communication Over Patient Portals

The health care professionals' experiences were also related to changes in communication over the patient portals. According to professionals, web-based patient-professional communication enhanced communication [39,42,43,46,48,49], but also changed the way of communicating [38,43,44,50], and sometimes led to deficiencies in communication [38,39,45,49].

Enhanced Communication

Several descriptive themes showed that web-based patient portals enhanced communication. When communicating on the internet, health care professionals provided patients with more direct answers [39,42,46,48,49] by keeping their writing short and concise [43,48]. Communicating on the internet also increased connectivity [48] and gave health care professionals an additional way to reach out to the patients [43]. Communication over the portal was more frequent, and occasionally [48], the patients' postings acted as triggers for further communication [43]. Sometimes, due to web-based communication, health care professionals were able to pay more focused attention to their patients' needs [48]. Communicating on the internet was perceived as easy because it enabled the professions to give direct answers to their patients without using intermediates, while also improving some patients' access to health services [43,48]. In addition, web-based communication was perceived to be safer than traditional face-to-face communication because portal communication allowed the patients to consider issues that had been discussed earlier [40,42] and it left *a trail* concerning the issues that had been discussed [42].

Change in the Means of Communication

Web-based patient-professional communication meant that health care professionals learned to communicate with patients in a new way [43,44]. When communicating on the internet, the dynamics of communication changed [50]. For example, there were no nonverbal clues [43]. Some health care professionals learned to communicate over the portal in a casual way, as they did on the internet in their personal lives [43], and they learned when to use web-based communication instead of other means of communication [38].

Deficiencies in Communication

Sometimes health care professionals were unable to communicate appropriately over the portal [38,39,45,49], and in the worst cases, the communication was insensitive [38], lacking in empathy, and overlooked the patients' cues of distress [39,40]. Insensitive communication did not encourage patients to attempt further communication [39], and in some cases, health care professionals mentioned that they selected a portal function to prevent patients from replying to message threads over the patient portal [45].

Sometimes communication was poor because health care professionals only had limited information available about the patients. For this reason, health care professionals were not always able to respond in a detailed manner, providing less specific advice to the patient [43]. In addition, health care professionals had variations in their expectations and attitudes toward portal communication. For example, although some professionals offered clear guidance over the portals, others were more equivocal [49].

Experiences Related to the Patients' Use of Patient Portals

Health care professionals also had experiences related to the patients' use of patient portals. Health care professionals were worried about the patients' interpretations and communication over the portal [39-41,43,44,46,47,49,50], but they also saw that web-based patient-professional communication had positive consequences for patients [39-43,46,48,50]. Furthermore, the possibility of communicating on the internet made health care professionals feel that the patients had high expectations from them [41,45,47,48].

Interpretation and Communication

Health care professionals were uncertain whether the patients understood the information that was communicated over the patient portals [43,46,50]. Some health care professionals reported that some of the patients were not able to understand the given information [41,44,47,50] and that the patients made their own interpretations [40]. Not understanding the given information left patients unsure about what to do next [39].

In some cases, patients used the portal inappropriately [49] and sent unsuitable messages to health care professionals [43,47,49]. For example, one patient sent a health care professional a picture from his vacation, which had nothing to do with the patient’s health status [49]. In addition, some patients provided insufficient [49] and poor-quality information [49,50]. For example, patients occasionally write lengthy descriptions devoid of clarity and clear questions [49].

Positive Consequences for Patients

From the viewpoint of health care professionals, web-based patient-professional communication had positive consequences for patients, such as better patient engagement [40,41,48], positive patient behavior [39], increased health status [46], and increased trust and satisfaction [48]. In addition, patients were better informed about their diseases and treatments [50] and started to follow the health care professionals’ recommendations more carefully [46]. As the patients had better knowledge, they were able to present more specific questions to health care professionals [43]. Occasionally, the level of their questions was so advanced that the health care professionals felt unprepared for them [43]. Portal communication also promoted patients’ involvement and participation, and the health care professionals asked for their patients’ opinions and concerns over the portal [39,42]. The threshold to discussing fearful or shameful topics was lowered as the negative feelings associated with these topics did not influence communication as much as they would in face-to-face settings [43,48].

Patients’ High Expectations of Professionals

When communicating over the portal, health care professionals reported that their patients expected rapid communication [41] and immediate responses to their electronic requests or status updates [45,47]. In addition, the professionals expressed concerns that patients often expected that professionals could view data, for example, on blood glucose levels, they entered, further influencing health care professionals’ endorsement of use [48].

Experiences Related to the Suitability of Patient Portals for Communication

The suitability of web-based patient-professional communication varied between patient cases. Patient portals were perceived as useful in some patients’ cases, such as for managing chronic conditions remotely [49] and learning about acute changes in their conditions [46]. However, as each patient’s case was unique [40], communicating over the portal was not the right choice for all cases. Health care professionals experienced difficulties with complex issues [38,50], complex patient cases [49,50], and complex information [41]. According to health care professionals, patient portals were not suitable for communication of complex issues [38,50], and occasionally, they felt that a richer communication platform was needed for more meaningful conversations [38], such as treatment and prognosis discussions [41]. Health care professionals prefer not to use patient portals for complex patient cases [49,50] due to limitations of the portals to convey deeper knowledge [43]. The professionals were also concerned about sharing complex information over the portal and agreed that sharing certain information, such as treatment discussions, was acceptable over the portal, whereas other types of information, such as prognoses, were not suitable for portal communication [41].

Experiences Related to the Convenience of Patient Portals for Communication

Some health care professionals appreciated the flexibility to answer portal messages when convenient [43,48,49] and saw asynchronous communication to be beneficial [43,46,48]. Sometimes, the professionals preferred the slower time line of web-based communication, as it allowed them to discuss and research their responses before sending an answer to the patients’ questions [45].

Not all health care professionals preferred web-based patient-professional communication. Some of the health care professionals saw portal communication as a nuisance [42] and felt that asynchrony disrupted care [38]. Poor convenience was also related to communication in those cases when health care professionals did not receive a confirmation that the patient had viewed a message [38]. This left the health care professionals unsure about whether the patients had read their instructions. Health care professionals also had differing views on the messaging functionality of patient portals [50]. For instance, nurses reported that they advised patients to ask questions over the portal, whereas some doctors preferred not to use the portal’s messaging function to answer questions [50].

Experiences Related to Changes in Roles

Web-based patient-professional communication has changed the roles of patients and health care professionals [50]. Some professionals reported that communicating with patients on the internet gave the patients more ownership and made them partners of the health care professionals [41,50], which again transformed the professionals’ roles [50]. However, despite the change in their role, the core element of the patient-professional relationship remained the same because after all, the health care professionals were still responsible for their patients [40,42].

Discussion

Principal Findings

This systematic review identified how health care professionals experienced web-based patient-professional communication via patient portals. A thematic synthesis produced 6 analytical themes that described health care professionals’ experiences concerning (1) health care professionals’ work, (2) changes in communication over the patient portal, (3) patients’ use of the patient portals, (4) the suitability of the patient portals, (5) the convenience of patient portals, and (6) changes in roles. The descriptive themes described in the analytical themes can be divided into positive and negative experiences, and some of the experiences seem to overlap so that even within the same study, some health care professionals seemed to experience the same feature positively, whereas others experienced it negatively.

Overall, health care professionals experienced web-based patient-professional communication positively, as it made their work more efficient, increased their knowledge about their patients’ situations, enhanced communication, had positive consequences for patients, and changed health care professionals' and patients' roles in a positive way. In addition, the findings of prior studies support some of the positive experiences found in this review [6,14,48]. In the review by Otte-Trojel et al [8], several studies mentioned how patient portals enabled building an ongoing, personal relationship that included mutual trust and responsibility. In this review, it seemed that web-based patient-professional communication is especially beneficial in that it produces positive consequences for patients, such as better patient engagement [35-39,42,44,46]. The findings of this review are in agreement with those of the earlier studies, which indicated that web-based communication may drive patient engagement to a new improved level [14,48,49].

In this review, health care professionals also reported that web-based patient-professional communication had a positive impact on health care professionals' and patients' roles [37,46] and that the patient-professional relationship remained the same as when having face-to-face contacts [40,42]. Prior studies have shown that health care professionals were worried that web-based communication may change patient-professional relationships [14,50]. In the study by Geerts et al [18], health care professionals expressed concerns that if they did not respond to a digital conversation quickly enough, the patient-professional relationship may change.

In this review, health care professionals' negative experiences were related to increased workload, time pressure, lack of expertise, communication problems, and the patients' interpretations and high expectations. In prior studies, the daily workload of health care professionals has also been reported to have a negative impact on web-based patient-professional communication [5,14,51]. In this review, health care professionals were worried about their patients' ability to understand information given over the patient portals. In the study by Baudendistel et al [52], health care professionals were likewise concerned about patients autonomously handling the information.

The analytical themes of (1) experiences of convenience and (2) changes in the means of communication included both positive and negative aspects. For example, some health care professionals appreciated the flexibility to communicate over the patient portal asynchronously [39,44,45], whereas some saw asynchronous web-based patient-professional communication as a nuisance [42]. In addition, some health care professionals felt that it was because of the portals that they had learned to communicate in a modern way, and they saw this as a positive thing [39,40]. However, some health care professionals also reported that the lack of nonverbal cues, such as body language, tone of voice, and gaze, made it difficult for them to assess whether the patient actually understood the information that they were providing [43].

Some of the included studies seemed to suggest an overall more positive insight into web-based patient-professional communication than other studies. Moreover, in some studies, health care professionals had overlapping positive and negative experiences about patient-professional communication. These kinds of overlapping experiences have also been detected in earlier studies [9,21]. There are a number of possible explanations as to why health care professionals in some studies have more positive experiences with web-based patient-professional communication than the others. First, within the same study, the differences might be explained by variations in the level of digital competence between health care professionals [53]. Those professionals who are more experienced in information technology usage might experience web-based patient-professional communication more positively because they know how to use patient portals and understand when to use them instead of other means of communication [38]. Then, professionals who are not very experienced users of patient portals might struggle with web-based communication and communicate in an inappropriate manner [34,35,41,45].

Second, differences in the experiences of health care professionals in different studies might be attributed to the differences in patient portals. Some of the portals might be less user-friendly than others. According to Kruse et al [25], patient portals obtain a higher level of acceptance if they are user-friendly. Alpert et al [38] noted that health care professionals were concerned about not receiving confirmations or notifications of whether their patients had checked their secure messages; thus, they had to contact patients by phone to ensure that the patients had received their messages. This indicates that the patient portal used in the study was not very user-friendly, as it eventually increased the health care professionals' workload and required an extra step in the care process [38].

Third, different health care organizations seem to have different practices in providing time and other resources for portal communication. In the study by Das et al [43], there was no time scheduled for web-based patient-professional communication; however, Nazi et al [48] noted that although health care professionals were worried about an increase in their workloads, it had been manageable because the use of the patient portal had grown in an organized way.

This systematic review showed that some health care professionals had learned how to deal with portal communications and identify when web-based patient-professional communication was appropriate [38]. However, some health care professionals seemed to struggle with web-based patient-professional communication, and this may even lead to feelings of fear and discomfort. In addition, a prior study reported that health care professionals found it difficult to learn new ways to communicate with patients over the patient portals [54], and it has also been reported that new skills are required to meet the new demands in the era of eHealth [55].

Health care professionals' web-based communication skills should be enforced by training because it has been proven to be effective in changing their behavior [14], such as improving the use of core communication skills [7,52] and enhancing empathic expression [56]. In addition, providing health care professionals with more training and technical support on patient portals might be beneficial to ensure positive staff attitudes [57]. More training for health care professionals on this topic might reduce their fear and discomfort concerning communicating over patient portals and assist them in communicating in an appropriate manner. In addition, training could possibly help health care professionals learn about new ways to counsel patients. In the study by Björk et al [58], physicians became more aware of how to communicate with their patients on the internet, how to simplify medical terms, and how to provide extensive medical information after using an Ask the Doctor service.

It appears that web-based patient-professional communication is not suitable for all kinds of communication, such as communicating about complex issues [34,46], complex patient cases [45,46], and complex information [41]. This kind of communication still requires face-to-face consultation or at least contact over the phone. However, in some cases, portal communication may be useful, and sometimes patients are even willing to pay for the possibility of contacting their health care professionals on the internet [59]. For example, web-based patient-professional communication has been seen to be useful in acutely changing conditions [46] and in cases where it is less fearful and shameful to discuss matters on the internet [39,44]. Currently, in the era of the COVID-19 pandemic, some patients display valiant acts of benevolence by preferring remote communication over face-to-face consultations to protect health care professionals from the virus [60]. Using telemedicine, such as patient portals, enables health care organizations to provide care and support for those who require it by minimizing the risk of exposure to patients and health care workers [60]. Due to the COVID-19 pandemic, the use of web-based patient-professional communication may increase considerably and quite rapidly.

Practical Implications and Further Research

As web-based patient-professional communication is becoming more common [14], more attention should be paid to it by policy makers, health care organizations, and educational institutions. Policy makers aiming for more patient-centered care should understand that developing health care professionals' web-based communication skills is essential for developing patient-centered health services, and attention should be paid to this issue in health programs and eHealth strategies. Health care organizations should invest in patient portals that are easy to use and functional. For example, they should provide notifications when patients read a message and ensure successful teamwork between professionals. In addition, having multimodal training materials available at sign-up and first portal log-in might be beneficial [61]. Moreover, organizations should schedule enough time for portal communication and arrange training on their use for health care professionals. According to a recent study by Hefner et al [61], training professionals could help them to communicate more efficiently on patient portals, for example, using secure messaging. Health care organizations and enterprises responsible for developing patient portals should also take into consideration the observations of health care professionals about the problems and benefits of patient portals. Health care professionals should also be included in the development of patient portals. Finally, educational institutes should take into account the increase in web-based patient-professional communication in their curricula.

As this review has shown, health care professionals have concerns over their patients' ability to understand the information provided over patient portals. In future research, it would also be interesting to examine whether patients agree with this concern. Due to the COVID-19 pandemic, the use of patient portals might increase, further transforming the nature of communication and making it an even more essential part of health care. Examining whether a change occurs in attitudes and experiences after the pandemic would also be interesting to provide some very current and up-to-date information on web-based patient-professional communication.

Limitations

The limitations of this review concern the search strategy, eligibility criteria, and heterogeneity of the selected studies. An electronic search of databases is effective, but it may not identify all eligible studies [62]. For example, only peer-reviewed studies were included in this review, and thus, relevant studies classified as gray literature might have been excluded [30]. This review only included studies that solely considered patient portals. Health care professionals also use other communication forms, such as video consultations [63], which were not considered in this review. Including studies examining video consultations might have provided more detailed information about remote communication; however, communication via video or text might also differ.

The studies included in this review originated from several different countries in which the level of digital health solutions usage varies. Most of the studies were conducted in the United States. Thus, the scope of the studies might not have provided generalized results about web-based patient-professional communication because health care systems and digital solutions vary between different countries.

Conclusions

Health care professionals had both positive and negative experiences related to web-based patient-professional communication. The positive experiences were most commonly related to the patients and patient outcomes, such as learning more about patients' situations and having better patient engagement. The negative experiences were related to aspects such as the additional workload on health care professionals, deficiencies in communication, patients' false interpretations, and the suitability of patient portals for communication. Negative experiences of health care professionals related to the use of patient portals seemed to be associated to the poor functionality of the portals and insufficient training and resourcing.

Acknowledgments

This work was supported by the Strategic Research Council at the Academy of Finland under grant number 327145.

Conflicts of Interest

None declared.

References

1. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: the 'black swan' for mental health care and a turning point for e-health. Internet Interv 2020 Apr;20:100317 [FREE Full text] [doi: 10.1016/j.invent.2020.100317] [Medline: 32289019]
2. From Innovation to Implementation. World Health Organization. 2016. URL: https://www.euro.who.int/__data/assets/pdf_file/0012/302331/From-Innovation-to-Implementation-eHealth-Report-EU.pdf [accessed 2020-11-25]
3. Weingart SN, Rind D, Tofias Z, Sands DZ. Who uses the patient internet portal? The PatientSite experience. J Am Med Inform Assoc 2006;13(1):91-95 [FREE Full text] [doi: 10.1197/jamia.M1833] [Medline: 16221943]
4. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
5. The Value of Personal Health Records and Web Portals to Engage Consumers and Improve Quality. Robert Wood Johnson Foundation. 2012. URL: https://www.rwjf.org/en/library/research/2012/07/the-value-of-personal-health-records-and-web-portals-to-engage-c.html [accessed 2020-11-25]
6. Sorensen L, Shaw R, Casey E. Patient portals: survey of nursing informaticists. Stud Health Technol Inform 2009;146:160-165. [Medline: 19592827]
7. Rigby M, Georgiou A, Hyppönen H, Ammenwerth E, de Keizer N, Magrabi F, et al. Patient portals as a means of information and communication technology support to patient- centric care coordination - the missing evidence and the challenges of evaluation. A joint contribution of IMIA WG EVAL and EFMI WG EVAL. Yearb Med Inform 2015 Aug 13;10(1):148-159 [FREE Full text] [doi: 10.15265/IY-2015-007] [Medline: 26123909]
8. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
9. Levinson W, Lesser CS, Epstein RM. Developing physician communication skills for patient-centered care. Health Aff (Millwood) 2010 Jul;29(7):1310-1318. [doi: 10.1377/hlthaff.2009.0450] [Medline: 20606179]
10. Shenson JA, Cronin RM, Davis SE, Chen Q, Jackson GP. Rapid growth in surgeons' use of secure messaging in a patient portal. Surg Endosc 2016 Apr;30(4):1432-1440 [FREE Full text] [doi: 10.1007/s00464-015-4347-y] [Medline: 26123340]
11. Reed M, Graetz I, Gordon N, Fung V. Patient-initiated e-mails to providers: associations with out-of-pocket visit costs, and impact on care-seeking and health. Am J Manag Care 2015 Dec 1;21(12):e632-e639 [FREE Full text] [Medline: 26760425]
12. Houston TK, Sands DZ, Jenckes MW, Ford DE. Experiences of patients who were early adopters of electronic communication with their physician: satisfaction, benefits, and concerns. Am J Manag Care 2004 Sep;10(9):601-608 [FREE Full text] [Medline: 15515992]
13. Jackson SL, DesRoches CM, Frosch DL, Peacock S, Oster NV, Elmore JG. Will use of patient portals help to educate and communicate with patients with diabetes? Patient Educ Couns 2018 May;101(5):956-959. [doi: 10.1016/j.pec.2017.11.004] [Medline: 29153758]
14. Alpert JM, Markham MJ, Bjarnadottir RI, Bylund CL. Twenty-first century bedside manner: exploring patient-centered communication in secure messaging with cancer patients. J Cancer Educ 2019 Jul 24:- epub ahead of print. [doi: 10.1007/s13187-019-01592-5] [Medline: 31342283]
15. Sieck CJ, Hefner JL, McAlearney AS. Improving the patient experience through patient portals: insights from experienced portal users. Patient Exp J 2018 Nov 6;5(3):47-54 [FREE Full text] [doi: 10.35680/2372-0247.1269]
16. Lindberg B, Nilsson C, Zotterman D, Söderberg S, Skär L. Using information and communication technology in home care for communication between patients, family members, and healthcare professionals: a systematic review. Int J Telemed Appl 2013;2013:461829 [FREE Full text] [doi: 10.1155/2013/461829] [Medline: 23690763]
17. Cornwall A, Moore S, Plant H. Embracing technology: patients', family members' and nurse specialists' experience of communicating using e-mail. Eur J Oncol Nurs 2008 Jul;12(3):198-208. [doi: 10.1016/j.ejon.2007.09.008] [Medline: 18406667]
18. Geerts PA, van der Weijden T, Loeffen PG, Janssen LE, Almekinders C, Wienhold TA, et al. Developing a patient portal for haematology patients requires involvement of all stakeholders and a customised design, tailored to the individual needs. BMC Med Inform Decis Mak 2019 Jul 11;19(1):129 [FREE Full text] [doi: 10.1186/s12911-019-0868-y] [Medline: 31296210]
19. Daniel F, Jabak S, Sasso R, Chamoun Y, Tamim H. Patient-physician communication in the era of mobile phones and social media apps: cross-sectional observational study on Lebanese physicians' perceptions and attitudes. JMIR Med Inform 2018 Apr 6;6(2):e18 [FREE Full text] [doi: 10.2196/medinform.8895] [Medline: 29625955]

20. Kujala S, Hörhammer I, Kaipio J, Heponiemi T. Health professionals' expectations of a national patient portal for self-management. Int J Med Inform 2018 Sep;117:82-87 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.005] [Medline: 30032968]
21. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
22. Lavender V, Gibson F, Brownsdon A, Fern L, Whelan J, Pearce S. Health professional perceptions of communicating with adolescents and young adults about bone cancer clinical trial participation. Support Care Cancer 2019 Feb;27(2):467-475. [doi: 10.1007/s00520-018-4337-4] [Medline: 29978324]
23. Feldman-Stewart D, Brundage MD, Tishelman C, SCRN Communication Team. A conceptual framework for patient-professional communication: an application to the cancer context. Psychooncology 2005 Oct;14(10):801-9; discussion 810. [doi: 10.1002/pon.950] [Medline: 16200514]
24. Sun S, Zhou X, Denny J, Rosenbloom T, Xu H. Understanding patient-provider communication entered via a patient portal system. Proc Am Soc Info Sci Tech 2013 Jan 24;49(1):1-4 [FREE Full text] [doi: 10.1002/meet.14504901387]
25. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
26. Ferreira A, Correia A, Silva A, Corte A, Pinto A, Saavedra A, et al. Why facilitate patient access to medical records. Stud Health Technol Inform 2007;127:77-90. [Medline: 17901601]
27. Voruganti T, Grunfeld E, Makuwaza T, Bender JL. Web-based tools for text-based patient-provider communication in chronic conditions: scoping review. J Med Internet Res 2017 Oct 27;19(10):e366 [FREE Full text] [doi: 10.2196/jmir.7987] [Medline: 29079552]
28. Wald JS. Variations in patient portal adoption in four primary care practices. AMIA Annu Symp Proc 2010 Nov 13;2010:837-841 [FREE Full text] [Medline: 21347096]
29. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
30. Aromataris E, Munn Z. JBI Systematic Reviews. The Joanna Briggs Institute. 2020. URL: https://synthesismanual.jbi.global [accessed 2020-11-25]
31. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
32. Polit DF, Beck C. Nursing research. In: Generating and Assessing Evidence for Nursing Practice. Ninth Edition. Philadeplhia, USA: Lippincott Williams & Wilkins; 2012.
33. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
34. Checklist for Qualitative Research. The Joanna Briggs Institute. 2016. URL: https://joannabriggs.org/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Qualitative_Research2017_0.pdf [accessed 2020-11-25]
35. Systematic Reviews of Qualitative Evidence. JBI Manual for Evidence Synthesis. 2020. URL: https://reviewersmanual.joannabriggs.org/ [accessed 2020-11-25] [WebCite Cache ID https://wiki.jbi.global/display/MANUAL/Chapter+2%3A+Systematic+reviews+of+qualitative+evidence]
36. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
37. Joseph-Williams N, Elwyn G, Edwards A. Knowledge is not power for patients: a systematic review and thematic synthesis of patient-reported barriers and facilitators to shared decision making. Patient Educ Couns 2014 Mar;94(3):291-309. [doi: 10.1016/j.pec.2013.10.031] [Medline: 24305642]
38. Alpert JM, Krist AH, Aycock RA, Kreps GL. Applying multiple methods to comprehensively evaluate a patient portal's effectiveness to convey information to patients. J Med Internet Res 2016 May 17;18(5):e112 [FREE Full text] [doi: 10.2196/jmir.5451] [Medline: 27188953]
39. Alpert JM, Dyer KE, Lafata JE. Patient-centered communication in digital medical encounters. Patient Educ Couns 2017 Oct;100(10):1852-1858 [FREE Full text] [doi: 10.1016/j.pec.2017.04.019] [Medline: 28522229]
40. Alpert JM, Morris BB, Thomson MD, Matin K, Brown RF. Implications of patient portal transparency in oncology: qualitative interview study on the experiences of patients, oncologists, and medical informaticists. JMIR Cancer 2018 Mar 26;4(1):e5 [FREE Full text] [doi: 10.2196/cancer.8993] [Medline: 29581090]
41. Alpert JM, Morris BB, Thomson MD, Matin K, Brown RF. Identifying how patient portals impact communication in oncology. Health Commun 2019 Nov;34(12):1395-1403. [doi: 10.1080/10410236.2018.1493418] [Medline: 29979886]
42. Bishop TF, Press MJ, Mendelsohn JL, Casalino LP. Electronic communication improves access, but barriers to its widespread adoption remain. Health Aff (Millwood) 2013 Aug;32(8):1361-1367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1151] [Medline: 23918479]

43. Das A, Faxvaag A, Svanæs D. The Impact of an eHealth Portal on Health Care Professionals' Interaction with Patients: Qualitative Study. J Med Internet Res 2015 Nov 24;17(11):e267 [FREE Full text] [doi: 10.2196/jmir.4950] [Medline: 26601678]
44. Elers P, Nelson F. Improving healthcare through digital connection? Findings from a qualitative study about patient portals in New Zealand. Aust J Prim Health 2018 Nov;24(5):404-408. [doi: 10.1071/PY17116] [Medline: 30149829]
45. Gerber DE, Beg MS, Duncan T, Gill M, Craddock Lee SJ. Oncology nursing perceptions of patient electronic portal use: a qualitative analysis. Oncol Nurs Forum 2017 Mar 1;44(2):165-170 [FREE Full text] [doi: 10.1188/17.ONF.165-170] [Medline: 28222081]
46. Kopanitsa G. A Qualitative Study of the Barriers and Opportunities for Adoption of Web-Portals for Doctors and Patients in Russia. J Med Syst 2017 Apr;41(4):62 [FREE Full text] [doi: 10.1007/s10916-017-0713-8] [Medline: 28271462]
47. Miller DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary care providers' views of patient portals: interview study of perceived benefits and consequences. J Med Internet Res 2016 Jan 15;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
48. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr 4;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
49. Sieck CJ, Hefner JL, Schnierle J, Florian H, Agarwal A, Rundell K, et al. The rules of engagement: perspectives on secure messaging from experienced ambulatory patient portal users. JMIR Med Inform 2017 Jul 4;5(3):e13 [FREE Full text] [doi: 10.2196/medinform.7516] [Medline: 28676467]
50. Vreugdenhil MM, Ranke S, de Man Y, Haan MM, Kool RB. Patient and health care provider experiences with a recently introduced patient portal in an academic hospital in the Netherlands: mixed methods study. J Med Internet Res 2019 Aug 20;21(8):13743 [FREE Full text] [doi: 10.2196/13743] [Medline: 31432782]
51. Yen P, McAlearney AS, Sieck CJ, Hefner JL, Huerta TR. Health information technology (HIT) adaptation: refocusing on the journey to successful hit implementation. JMIR Med Inform 2017 Sep 7;5(3):e28 [FREE Full text] [doi: 10.2196/medinform.7476] [Medline: 28882812]
52. Baudendistel I, Winkler EC, Kamradt M, Brophy S, Längst G, Eckrich F, et al. Cross-sectoral cancer care: views from patients and health care professionals regarding a personal electronic health record. Eur J Cancer Care (Engl) 2017 Mar;26(2):-. [doi: 10.1111/ecc.12429] [Medline: 26840784]
53. Koivunen M, Saranto K. Nursing professionals' experiences of the facilitators and barriers to the use of telehealth applications: a systematic review of qualitative studies. Scand J Caring Sci 2018 Mar;32(1):24-44. [doi: 10.1111/scs.12445] [Medline: 28771752]
54. Varsi C, Ekstedt M, Gammon D, Børøsund E, Ruland CM. Middle managers' experiences and role in implementing an interactive tailored patient assessment ehealth intervention in clinical practice. Comput Inform Nurs 2015 Jun;33(6):249-257. [doi: 10.1097/CIN.0000000000000158] [Medline: 25988851]
55. Öberg U, Orre CJ, Isaksson U, Schimmer R, Larsson H, Hörnsten A. Swedish primary healthcare nurses' perceptions of using digital eHealth services in support of patient self-management. Scand J Caring Sci 2018 Jun;32(2):961-970. [doi: 10.1111/scs.12534] [Medline: 28960451]
56. Bonvicini KA, Perlin MJ, Bylund CL, Carroll G, Rouse RA, Goldstein MG. Impact of communication training on physician expression of empathy in patient encounters. Patient Educ Couns 2009 Apr;75(1):3-10. [doi: 10.1016/j.pec.2008.09.007] [Medline: 19081704]
57. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
58. Björk AB, Hillborg H, Augutis M, Umefjord G. Evolving techniques in text-based medical consultation - physicians' long-term experiences at an Ask the doctor service. Int J Med Inform 2017 Sep;105:83-88. [doi: 10.1016/j.ijmedinf.2017.05.013] [Medline: 28750914]
59. Virji A, Yarnall KS, Krause KM, Pollak KI, Scannell MA, Gradison M, et al. Use of email in a family practice setting: opportunities and challenges in patient- and physician-initiated communication. BMC Med 2006 Aug 15;4:18 [FREE Full text] [doi: 10.1186/1741-7015-4-18] [Medline: 16911780]
60. Kavoor AR, Chakravarthy K, John T. Remote consultations in the era of COVID-19 pandemic: preliminary experience in a regional Australian public acute mental health care setting. Asian J Psychiatr 2020 Jun;51:102074 [FREE Full text] [doi: 10.1016/j.ajp.2020.102074] [Medline: 32294583]
61. Hefner JL, Sieck CJ, Walker DM. Patient and physician perspectives on training to improve communication through secure messaging: Clarifying the rules of engagement. Health Care Manage Rev 2020 May 05:- epub ahead of print. [doi: 10.1097/HMR.0000000000000279] [Medline: 32379081]
62. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]

63. Shaw SE, Seuren LM, Wherton J, Cameron D, A'Court C, Vijayaraghavan S, et al. Video consultations between patients and clinicians in diabetes, cancer, and heart failure services: linguistic ethnographic study of video-mediated interaction. J Med Internet Res 2020 May 11;22(5):e18378 [FREE Full text] [doi: 10.2196/18378] [Medline: 32391799]

Abbreviations

EHRs: electronic health records
JBI: Joanna Briggs Institute
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

---

Edited by G Eysenbach; submitted 19.06.20; peer-reviewed by J Alpert, J Hefner; comments to author 28.10.20; revised version received 10.11.20; accepted 14.11.20; published 08.12.20.

Please cite as:
Laukka E, Huhtakangas M, Heponiemi T, Kujala S, Kaihlanen AM, Gluschkoff K, Kanste O
Health Care Professionals' Experiences of Patient-Professional Communication Over Patient Portals: Systematic Review of Qualitative Studies
J Med Internet Res 2020;22(12):e21623
URL: http://www.jmir.org/2020/12/e21623/
doi:10.2196/21623
PMID:33289674

Review

Capturing the Impact of Patient Portals Based on the Quadruple Aim and Benefits Evaluation Frameworks: Scoping Review

Melita Avdagovska[1], PhD; Devidas Menon[1], PhD; Tania Stafinski[1], PhD

School of Public Health, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Melita Avdagovska, PhD
School of Public Health
University of Alberta
3021, Research Transition Facility 8308 114 Street
Edmonton, AB, T6G 2V2
Canada
Phone: 1 7809083334
Email: avdagovs@ualberta.ca

Abstract

Background: Despite extensive and continuing research in the area of patient portals, measuring the impact of patient portals remains a convoluted process.

Objective: This study aims to explore what is known about patient portal evaluations and to provide recommendations for future endeavors. The focus is on mapping the measures used to assess the impact of patient portals on the dimensions of the Quadruple Aim (QA) framework and the Canada Health Infoway's Benefits Evaluation (BE) framework.

Methods: A scoping review was conducted using the methodological framework of Arksey and O'Malley. Reporting was guided by the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) extension for scoping reviews. A systematic and comprehensive search was conducted using the Ovid platform, and the following databases were searched: Ovid MEDLINE (R) ALL (including epub ahead of print, in-process, and other nonindexed citations), EMBASE, and PsycINFO. CINAHL on the EBSCO platform and Web of Science were searched for studies published between March 2015 and June 2020. A systematic gray literature search was conducted using the Google search engine. Extracted data were tabulated based on a coding template developed to categorize the literature into themes and areas of interest.

Results: A total of 96 studies were included for data extraction. The studies were categorized based on the QA dimensions, with strict adherence to the definitions for each dimension. From the patients' perspective, it was determined that most evaluations focused on benefits and barriers to access, access to test results, medication adherence, condition management, medical notes, and secure messaging. From the population perspective, the evaluations focused on the increase in population outreach, decrease in disparities related to access to care services, and improvement in quality of care. From the health care workforce perspective, the evaluations focused on the impact of patients accessing medical records, impact on workflow, impact of bidirectional secure messaging, and virtual care. From the health system perspective, the evaluations focused on decreases in no-show appointments, impact on office visits and telephone calls, impact on admission and readmission rates and emergency department visits, and impact on health care use. Overall, 77 peer-reviewed studies were mapped on the expanded version of the BE framework. The mapping was performed using subdimensions to create a more precise representation of the areas that are currently explored when studying patient portals. Most of the studies evaluated more than one subdimension.

Conclusions: The QA and BE frameworks provide guidance in identifying gaps in the current literature by providing a way to show how an impact was assessed. This study highlights the need to appropriately plan how the impact will be assessed and how the findings will be translated into effective adaptations.

(J Med Internet Res 2020;22(12):e24568) doi:10.2196/24568

KEYWORDS

patient portals; quadruple aim; benefits evaluation framework; mobile phone

Introduction

Background

Electronic patient portals are secure websites tied to an institutional electronic health record (EHR) system from which patients can view their medical information. These types of records are populated with a person's lifetime health history. The information comes from various sources, including community clinics, hospitals, physicians, pharmacies, and laboratories. Many allow patients to view appointments, medical test results, and medication therapies and communicate with their health care providers through a secure platform [1,2]. Although evidence suggests that the implementation of patient portals can have a positive impact on patient care and patient outcomes, many health systems have been slow to adopt them because of various concerns [3-6].

Patients have expressed concerns about accessing and maintaining health information in a private and secure manner [2,7,8]. As these portals are web-based tools, some worry that their data could be seen by other individuals and by insurance companies [9]. Others have expressed concerns around optimal design and functionality [2]. Furthermore, not all patients have access to a computer, smartphone, or tablet to access their record [10].

Health care providers have conveyed concerns about the implementation, availability, and impact of portals on the patient-provider relationship. Physicians continue to be concerned that portals will increase their workload, without a mechanism for remuneration in fee-for-service models [11]. In addition, there are some uncertainties around physician obligations with respect to portal use [9,11-14]. Providers are concerned that this type of health technology will start replacing office visits and thus have an impact on the way care is provided [15]. Another practice-related concern is the absence of transparency surrounding the provider's role and accountability with respect to patient portals and protection of patient data [3].

For governments, the challenge of implementing portals has been the upfront cost of establishing an effective and efficient system [14,16,17]. Furthermore, studies show that the majority of health information technology investments are struggling to achieve the anticipated benefits [18-21].

Despite continuing extensive research in this area, the implementation and adoption of these systems remains a convoluted process. First, although various reviews [22-31] have been conducted in this area recently, none have focused on the impact of patient portals within the context of the 4 specific dimensions of the Quadruple Aim (QA) framework [32]. Second, the majority of published reviews have examined one characteristic, such as engagement, barriers and facilitators, outcomes, or communication, and these reviews usually emphasize patients or health care providers. Third, no reviews have looked at which patient portal functions are most commonly used to evaluate impact based on the Canada Health Infoway's Benefits Evaluation (BE) framework [33].

Two frameworks for analysis were incorporated: the QA and BE frameworks. The QA framework is a modification of the established Triple Aim Framework [34] of health care improvement, which focuses on evaluating 3 dimensions of care: improving the health of populations, improving the patient and caregiver experience, and reducing the per capita cost of health care. The QA framework [32] focuses on improving the work life of providers as the fourth dimension, providing a more comprehensive approach to the evaluation of health technology. Definitions for each of the QA dimensions were used to identify measures as they related to what is considered significant to the patient (ie, preferences, satisfaction, communication, access, engagement, use, etc), population (ie, equity, access, disparities, etc), health system (ie, costs, utilization, etc), and health workforce (ie, satisfaction, workload, preferences, etc).

The BE framework [33] as described and expanded by Lau et al [35] was used to organize measures from peer-reviewed studies. The BE framework was introduced in 2006 by Canada Health Infoway to determine how evaluations might be conducted to capture and measure relevant indicators. The indicators are divided into 8 categories (system quality, information quality, service quality, quality, access, productivity, use, and use satisfaction) and 20 subcategories. Lau et al [35], in their review of systematic reviews, added patient and provider, implementation, and change/improvement as additional categories, which were incorporated in this study.

Objectives

The purpose of this study is to explore what is known about patient portal evaluations and provide recommendations for future endeavors. It specifically addresses the following research questions:

1. How is the impact of patient portals measured from the standpoint of the 4 specific dimensions (patients, population, health care workforce, and health system) of the QA framework?
2. What components from the BE framework (as expanded by Lau et al [35]) are most commonly evaluated to measure impact?

Methods

Methodology

A scoping review was conducted following the 5 steps identified by Arksey and O'Malley [36]: identifying the relevant research question; identifying the various relevant studies in this field; selecting studies; charting the data; and collating, summarizing, and reporting the results. This type of review is recognized as particularly useful for exploring topics with inconsistencies in the current evidence, as it appropriately captures broad and ambiguous topics and approaches.

In this study, *impact* is defined as the *overall effects, direct or indirect, of a policy, strategy, program or project* (in this case, patient portals) [37].

No ethics approval was sought or required for this study, as it did not involve any human subject because it was only focused on reviewing the literature.

Data Sources and Searches

To capture the wide array of studies that may be relevant to this topic, all study designs were included. A gray literature search was developed to capture all relevant publications, such as government and evaluation reports. Publications that study the same intervention in the same set of patients were matched and classified as a single study.

The inclusion and the exclusion criteria are described in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Studies with any defined impact and outcomes of tethered patient portals or personal health records
- Studies with relevant impact and outcomes of tethered patient portals or personal health records

Exclusion criteria

- Studies without any defined impact and outcomes of tethered patient portals or personal health records
- Studies with no relevant impact and outcomes of tethered patient portals or personal health records
- Studies describing impact and outcomes of untethered patient portals or personal health records
- Non-English language
- Documents published before 2015
- Abstracts
- Commentaries
- Opinions
- Articles summarizing study findings
- Clinical trials and clinical trial recruitment

With support from an experienced medical information specialist, a search strategy for peer-reviewed papers was developed and tested through an iterative process. Another senior information specialist peer reviewed the strategies before execution using the peer review of electronic search strategies (PRESS) checklist [38]. The following databases were searched using the OVID platform: Ovid MEDLINE, including epub ahead of print, in-process, and other nonindexed citations, EMBASE, and PsycINFO. CINAHL (Cumulative Index of Nursing and Allied Health Literature) on the EBSCO platform and Web of Science were also searched. All searches were performed on June 8, 2020. Strategies used a combination of controlled vocabulary (eg, *Patient Portals*, *Electronic Health Records*, *Patient Access to Records*) and keywords (eg, *health portal*, *EHR portal*, *ehealth patient access*). Vocabulary and syntax were adjusted across databases. Specific details regarding the strategies appear in Multimedia Appendix 1. After removal of all duplicates, the total number of articles remaining was 34,128. Citations retrieved via the searches of electronic databases were imported to Covidence, a Cochrane-supported software designed for conducting reviews.

A systematic gray literature search was conducted using the Google search engine in Edmonton, Alberta, Canada, between February 13 and 25, 2020. The search term *patient portal* was combined with the terms *impact* or *outcome*. The first 100 hits were considered from each combination. In addition, organizational websites of Canada Health Infoway, Canadian Agency for Drugs and Technologies in Health, the National Institute for Health and Care Excellence, the International Network of Agencies for Health Technology Assessment, and the World Health Organization were scanned. Finally, the reference lists of the included articles were searched manually.

Study Selection

The relevance of the retrieved studies was assessed using the inclusion criteria to ensure that they were related to the topic of this study. All citations were reviewed by titles and abstracts. All articles that focused on topics other than patient portals or personal health records were eliminated. Thus, 2259 articles remained, the titles and abstracts of which were screened independently by 3 researchers (MA, TS, and DM) who applied the inclusion and exclusion criteria. For quality assurance, 9.96% (225/2259) of the articles were reviewed by more than one researcher. No significant discrepancies were noted.

Potentially relevant citations were then retrieved and divided among the 3 researchers for screening using the same inclusion and exclusion criteria. For quality assurance, 10% (9/96) of the papers were reviewed by more than one researcher. No significant discrepancies were noted.

Studies that focused on *untethered* patient portals or personal health records, which were not available in English or were conference abstracts, unpublished dissertations, opinions, or editorials, were excluded.

A total of 10 reviews [22-31] that fit the inclusion criteria were identified. References from each of the reviews were scanned, and 58 articles that met the inclusion criteria were identified. Of these, 7 were reviewed and determined to be captured in the studies already included.

Data Extraction

As is customary in scoping reviews, an iterative approach was used to extract data from the selected studies. A data extraction form was developed and reviewed to categorize the literature into themes and areas of interest, which varied by study type. The following elements were considered, discussed, and incorporated in the data extraction form (tabular format): authors, title, publication date, country, type of source, study setting, research questions, aims, data collection methods, vendor, patient portal name, patient portal functions, deployment date, end date of the project, type of evaluation, number of patients impacted, number of staff impacted, intervention, control, length of follow-up, benefit area, net benefit indicators, measures, definition of measures, tools used to measure, results and key themes identified in the study, identified success factors, identified challenge factors, identified recommendations, and other considerations, thoughts, and notes.

The extraction form was piloted with several sources and any identified issues were corrected.

Quality Assessment

As scoping reviews include a broad range of information sources and topics, no critical appraisal of the quality of the included papers was conducted. Although there are various suggested [39] approaches for accomplishing this, there is no consensus among experts in the field on this matter.

Data Analysis and Synthesis

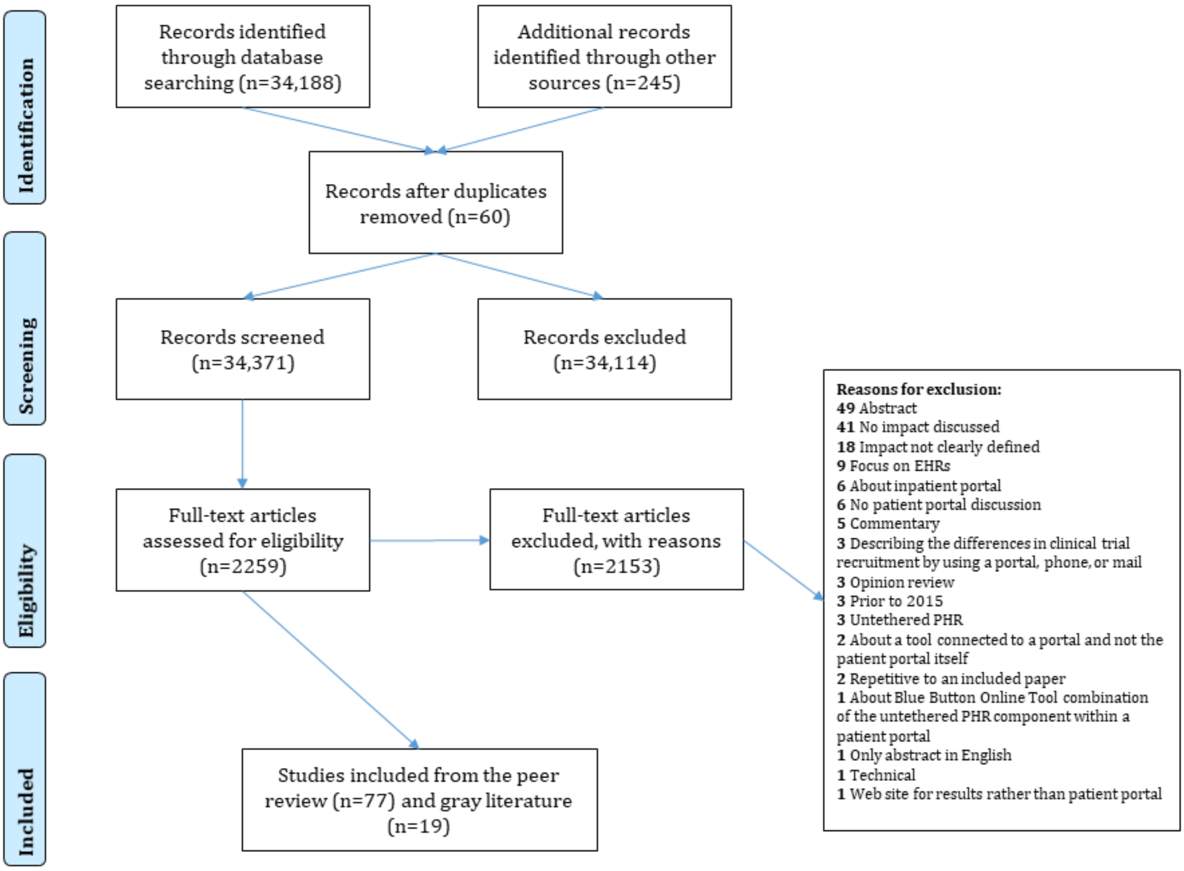
Extracted data were tabulated based on a coding template developed to categorize the literature into themes and areas of interest, which varied by study type and QA perspective. The studies were categorized based on the QA dimensions, with strict adherence to the definitions for each dimension [32,34]. Furthermore, the evaluated patient portal functions were mapped onto the BE framework [33,35]. A descriptive, analytical approach was used to summarize the outcomes of the studies. In addition, a list of the various terms/outcomes/variables that were used to describe *impact* was compiled.

Results

Results of the Literature Search

A total of 34,371 citations were identified through the peer-reviewed and gray literature searches. From these, 241 citations were considered for a full-text review. In total, 96 studies were included for data extraction. The search strategy results are described using the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram, as shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram. EHR: electronic health record; PHR: personal health record.



Overall Description of Included Studies

The 96 included studies employed a variety of methods (mixed, n=21; prospective, n=8; qualitative, n=12; randomized controlled trial [RCT], n=3; retrospective, n=25; and survey, n=27). They were conducted between 2015 and 2020 (2015, n=21; 2016, n=25; 2017, n=14; 2018, n=15; 2019, n=13; and 2020, n=8) in Australia (n=2), Canada (n=21), China (n=1), Finland (n=1), the Netherlands (n=2), Norway (n=1), Spain (n=1), Sweden (n=6), the United Kingdom (n=1), and the United States (n=60). Although not all studies specified a clinical problem, most were related to cardiology procedures and conditions, depression, posttraumatic stress disorder, HIV, substance use disorder, anxiety, schizophrenia, neurological issues, pregnancy, and diabetes. A summary of the included studies is captured in Multimedia Appendix 2 [27,40-136].

Overall, 32 peer-reviewed studies [40-71] evaluated the patient portal in general, with all available functions. Secure messaging and medical notes (OpenNotes) were the most commonly evaluated individual portal functions (11 studies each). Five studies assessed multiple functions, such as secure messaging and refills or secure messaging and medication reminders. Gray literature studies evaluated the patient portal, in general, as their focus was on appraising various identified net benefit areas (ie, quality, access, system use, etc), and patient and provider satisfaction with the available functions.

QA Dimensions

The following sections summarize the studies according to the QA dimensions (Multimedia Appendices 3-6) [27,43,44,46-48, 53,55-58,60-63,65,67,68,72-103,105-114,119,126,127,129].

Patient Perspective

The patient perspective was explored in 44 peer-reviewed [43,44,46-48,53,55,56,58,60-63,65,67,68,72-99] and 15 gray literature [100-114] studies. Several methods (mixed, n=12; observational, n=5; qualitative, n=5; RCT, n=1; retrospective, n=14; and survey, n=22) were applied to gain insights from patients through surveys, interviews, focus groups, and administrative data. Of the 59 studies, 35 were from the United States, 17 from Canada, 2 from the Netherlands, and 1 each from Australia, China, Norway, Sweden, and the United Kingdom. The studies usually explored the impact of the patient portal in general, with only a few focusing on various portal functions, such as test results, medical notes (OpenNotes), secure messaging, or prescription refills.

Benefits of Access

Many of the studies [47,68,77,80,82,100,101,109-112,114] have explored patient experiences with access to patient portals and subsequently access to their own medical information. All of the studies reported that users were highly satisfied with the access, and in one study, 97% of survey respondents stated that they would definitely or probably recommend the portal to other clients and families [110].

Moll et al [58] observed that patients considered access to information as a means of patient empowerment and involvement. This was also observed by Crouch et al [43], who found that the use of the portal was associated with significantly higher levels of patient activation and levels of patient satisfaction around timely appointments, care, and information. In addition, studies found that if the health care provider encouraged access, the likelihood of patient enrollment and use of the service was much greater [47,62,73].

Reed et al [60] found that 9 of 10 patients believed the portal improved their health care convenience. In addition, access to the information allowed for better engagement of patients with providers, as they had more knowledge about their health [61]. Furthermore, the information eliminated the time pressure felt during short appointments [60]. A different study reported that 28% of patients/families avoided making a telephone call to a health care provider because they could access health information electronically [111]. Convenience was also noted in the findings of another study, in which 27.2% of patients reported savings in terms of time to travel, time off work, gas, and parking [63]. In a study by Graham et al [48], 48% of users reported avoiding a clinic visit and 2.7% avoided an emergency department visit. Convenience was also described because of the ability to make web-based appointments rather than by calling the office [106]. In another study, 27.4% of patients indicated that they had used the patient portal at least once to request an appointment with a primary care provider rather than making a telephone call [114].

Access to patient portal information decreased stress levels because of appointment preparedness [114]. One study reported that 40% of respondents found the portal useful, as it allowed them to plan and follow up on upcoming appointments [67]. In another study, 60% of respondents felt the portal had resulted in an increased sense of partnership with their health care provider, compared with 50% of respondents who felt the portal had positively impacted their relationship with their health care provider [109].

Although most of the studies reported a positive impact because of portal access, one study found little evidence that the portal led to feelings of greater involvement in the care process, improved ability to express concerns to providers or enhanced relationships with providers, or reduced number of in-person visits [55].

Barriers to Access

Several studies described patient-reported barriers to access to patient portals [67,73,108]. These barriers were related to privacy, security, and technical difficulties when patients attempted to enroll or use the patient portal. Giardina et al [73] found that 52.6% of the participants wanted portal improvements in terms of display, usability, and notifications. Approximately 24% of patients had higher expectations based on their idea of what functionalities a patient portal should provide, whereas 22% experienced usability problems [67]. Another study [108] found that low user adoption was because of technical issues experienced by patients during enrollment. Patients were unwilling to spend extra time to find solutions and eventually abandoned the creation of an account [108].

Access to Test Results

The most commonly used portal function was the access to laboratory or diagnostic test results. The studies assessing this function concluded that the impact was multifaceted, providing patients with convenience, knowledge, tracking of information, decreased anxiety, and the need for fewer appointments [58,61,68,73,77,83,97,103,106,113]. Visual indications were used to determine whether the test results were normal or abnormal (ie, green or red color). Patients described laboratory results as the most important information for them to access. Getting real-time information of laboratory tests before appointments led to increased awareness about personal health. A study found that the availability of web-based radiology reports was associated with increased patient use of the system, with a likelihood ratio of 2.63 [97]. The rates of laboratory test–related anxiety were low. Another study found that 68.41% wanted access to new information on the same day or after a day, whereas the remaining patients were willing to wait anytime between 2 weeks and 1 month, depending on the type of test [58].

Although access to test results was described as the most appreciated function, several studies found that it led to concerns. Two studies [73,77] addressed the concern of the inability of patients to completely understand the laboratory or diagnostic test results in their medical records. Both studies found that patients did not feel that health care providers gave sufficient information when commenting on results. The study by Giardina et al [73] showed that 63.2% of the participants reported that their physician did not include a note explaining the result. Most often, the medical terminology used to describe the test results led to the inability to interpret if they were normal or abnormal. The problem of not understanding results led to apprehension and anxiety until the patient was able to connect with their provider and obtain clarification [73,77]. A study found that one of every 6 patients who underwent magnetic resonance imaging or computerized tomography scans reported a clear understanding of their results when first receiving them through the portal [83]. Patients wanted to receive all their results, even the abnormal ones, but they needed more timely notifications and guidance by their provider in interpreting them [90]. However, some patients preferred to have the potentially concerning test results verbally communicated by a health care professional [90].

Medication Adherence

Several studies have explored the correlation between patient portal use and medication adherence because of web-based reminders for refills and requests for prescription renewals [44,81,85,86,89,94,106,109,112]. A study found that once new users were given mobile access to the portal, there was a statistically significant improvement in adherence to oral diabetes drugs and lower glycemic levels [81]. These improvements were greater among patients with a higher clinical need at baseline (glycated hemoglobin [HbA_{1c}] level >8%) and more modest but still statistically significantly better among patients with lower initial glycemic levels [81]. Wright et al [86] found that adherence to antihypertensive medications increased if patients had access to their progress notes. The secure messaging function had a similar effect on the likelihood of achieving HbA_{1c} control, as patients who only read email also had significantly lower mean HbA_{1c} values than that of nonusers [44]. Similarly, another study observed that secure messages had the greatest impact on diabetes medical management considerations in terms of HbA_{1c} test completed or missed therapy intervention [96].

One study found small, statistically significant, meaningful improvements in physiological measures among patients with diabetes who initiated and sustained the use of refills through the patient portal [94]. The refill function, in combination with secure messaging, had a greater impact on HbA_{1c} levels. Another study observed stable refill adherence over time among portal users compared with small declines among nonusers [85]. Satisfaction with the refill portal function was high, as 69% would recommend e-refill requests to other patients, family, or friends, and 63% would request all or most of their prescription refills electronically [112]. Furthermore, a nation-wide survey in Canada found that when prescriptions were lost or damaged, 17% of patients decided to go without the medication [106]. Consequently, portals were determined to be effective as a tool to update medication lists and had the potential to augment the existing phone-based medication update process [89].

Condition Management

Patients described portal access as a way to monitor their conditions and be more proactive in their care. The severity of the disease predisposed the level of use [62]. A study that measured the acceptability and clinical outcomes of the portal in parents of children with moderate or severe asthma observed that parents used the portal as a decision-support tool that allowed for improved knowledge about the condition [46]. The more severe the child's condition, the higher the acceptance and use of the portal [46]. Crouch et al [43] concluded that higher portal use was associated with positive clinical and behavioral characteristics related to the management of chronic conditions. A study found that access to the portal added value in the received care during pregnancies [47]. Broman et al [87] found that portal use was effective in postoperative care and follow-up. Another study reported that 88% of survey respondents reported that portal access allowed for better health management [109].

However, a few studies found that portal use did not enhance patients' experiences. Two-thirds of persistent users responded that they did not feel that the portal supports them in most lifestyle choices [62]. A study observed that portal use among patients with chronic conditions enrolled in a care coordination program did not demonstrate a statistically significant improvement in self-efficacy and perception of health status [65].

Medical Notes

Access to medical notes (usually referred to as OpenNotes in the literature) through patient portals was another component of several studies. A study found that almost all patients described enhanced comprehension about their disease and care because of access to clinicians' notes, as the notes refreshed their memory and clarified their understanding of visits [74]. Patients reported that the medical notes eased their uncertainty, relieved anxiety, and facilitated control [74]. Denneson et al [75] found that reading OpenNotes helped 49% of patients have feelings ranging from very to extremely in control of their health care. Another study observed that access to notes increased patient trust toward their health care providers [92]. Notes not only provided a way for patients to learn about their condition but also checked for any inaccuracies and made face-to-face time more effective [72].

Higher levels of reading notes were associated with higher shared decision-making levels [78]. A study observed that patients who read >4 notes were 15% more likely to have high scores for clinician effort in helping them understand health issues and 16% more likely for clinician efforts to include them in the plan of care [78]. The study concluded that there was a strong correlation between shared decision making and the transparency provided by OpenNotes. A similar finding was observed by Walker et al [79], who found that transparency through notes helped patients feel more engaged in their care.

Caregivers found access to clinicians' notes valuable. A study found that 55% of caregivers reported reading notes helped them remember to get the patient's tests done, and 92.3% reported reading notes helped them understand the reason for the patient's referral to a specialist [76]. The same study found that caregiver access to notes had little to no negative impact on caregiver-provider relationships [76]. Wolff et al [98] found that 35.5% of caregivers viewed doctor notes because they were unable to visit.

For subsets of patients, access to medical notes increased their anxiety levels [74]. One study found that 26% of the patents experienced stress or worry sometimes, whereas 8% reported often or always [75]. The study also reported that 18% of patients felt upset sometimes after reading their notes, compared with 8% who reported often or always. Furthermore, race and ethnicity affected the levels of access to the notes. Minorities and patients with a lower socioeconomic status accessed notes at lower rates than patients who were White and had a high socioeconomic status [93].

Secure Messaging

The secure messaging function was most commonly used to request clarification, ask condition-related questions, or inform providers or patients about any health changes [99]. Secure messages were described as a tool to recognize and decrease any gaps in care [96]. A study found that secure messaging allowed for efficient bidirectional radiologist-patient communication [97]. Haun et al [88] noted that the majority of the respondents used secure messaging at least once a year, and less than 15% reported never using secure messaging. The same study observed that patients were satisfied with secure messaging, as it provided a safe and secure communication tool that was easy to use and saved time [88]. Another study found that patients reliably read messages sent by their physicians, and the rate of unread messages was 3.1% at 21 days [84]. Furthermore, secure messaging improved the management of clinical outcomes. Petullo et al [95] found that active secure messaging use was associated with a 0.156% lower HbA_{1c} compared with inactive patients ($P<.001$) and a 0.263% lower HbA_{1c} compared with active nonusers ($P<.001$). Similar rates were observed by Devkota et al [44], in which patients who read and wrote emails had significantly ($P<.001$) lower average HbA_{1c} values compared with nonusers. A study observed that patients who used the portal, compared with nonusers, were 24% more likely to achieve blood pressure control; however, after adjusting for sociodemographic factors, this association was no longer present because of low rates of portal use among minorities and disadvantaged patients [56].

The main barrier to the use of secure messaging was the unresponsiveness of health care providers to the messages sent by patients, which led to increased rates of telephone calls [67].

Population Perspective

Enhancing population health through decreasing disparities and elevating access to needed health services was explored in 5 peer-reviewed [52,115-118] and 8 gray literature [100,103,106,109,112-114,119] studies. The outcomes were evaluated through various methods, including RCTs (n=2), retrospective observational studies (n=2), qualitative studies (n=1), mixed studies (n=6), and cross-sectional surveys (n=2). One study was from Australia, 5 were from the United States, and 7 were from Canada. In addition to EHR and portal data, surveys, interviews, and focus groups were the most common sources of data. The studies analyzed the capacity of patient portals to increase vaccination rates, equity in access to timely care, and population empowerment.

Increase in Population Outreach

The RCTs examined the effectiveness of patient portals in improving influenza vaccination rates [115,117]. Although influenza infections have the potential to lead to serious health issues and increased access to health care services, vaccination rates continue to be low, necessitating the need for innovative outreach interventions to remind and encourage citizens to get the shot. As EHRs deliver real-time data identification, tethered patient portals were seen to have the potential to identify unvaccinated populations and enable implementation of portal-based cost-effective interventions. Cutrona et al [115] found a small but statistically significant improvement in the completion of influenza vaccination among portal users, especially by patients who opened reminder messages sent through the portal. Although a very small proportion (0.3%) of patients accessed the various influenza educational materials, Szilagyi et al [117] established a correlation between the higher numbers of reminders that led to higher vaccination rates by portal users. The portal reminders had a small, statistically significant effect on increasing rates among adults aged from 18 to 64 years, male patients, non-Hispanic patients, and those not vaccinated in the previous 2 years.

Decrease in Disparities Related to Access of Care Services

Foster et al [118] found that there were existing disparities between patient groups related to health care information access in emergency departments. African Americans and Hispanics had the lowest portal use rates, which led to disparities in medical information access. In their retrospective study, Lyles et al [116] reported a significant improvement in statin adherence regardless of race and ethnicity once patients increased portal use. The authors concluded that portal use could improve various health behaviors. Similarly, another study found that because of the ability to request and receive prescription renewals through a portal, patients did not need to travel, arrange care, or take time off work, which increased medication adherence and decreased wait time (74%) [112]. Another study concluded that *if only affluent, well-educated patients can access portals and understand them, then these technologies could potentially worsen health disparities* as one of the factors contributing to disparities in access were the decisions by providers to selectively offer access [100].

Improvement in Quality of Care

Two studies captured experiences of users who acknowledged that portals improved their quality of care and the ability to manage care because of information access [52,109]. However, because of low uptake by health care providers, they were uninformed about the portal. Two additional studies reported that Canadians felt more engaged and active, as the portal allowed them to have more informed discussions with their doctor [106,119]. Two evaluations found that access to health information contributed to easier access to services and acted as an *expansion of the standard 15 min consultation appointment* [113,114]. These studies demonstrated the readiness and willingness of patients to be more engaged in their health care. However, some apprehension was experienced, which could be elucidated by the *empowerment effect* related to web-based access to results and related information [103].

Health Care Workforce

In total, 18 studies [40-42,50,51,54,55,57,64,66,82,87,120-125] and 3 reports [114,126,127] addressed the health care workforce perspective through various methods (mixed methods, n=7; prospective observational, n=2; qualitative, n=6; and surveys, n=6), with a focus on conducting formative and process evaluations. A total of 11 studies were conducted in the United States, 5 in Sweden, 4 in Canada, and 1 in China. They explored health care provider experience with patient access to medical records, laboratory and diagnostic results, secure messaging, and uploading of images and symptoms. To gather data, focus groups, semistructured interviews, and surveys were used. Many of the studies used a combination of different data collection methods.

Impact of Patients Accessing Medical Records

Although health care providers generally agreed with the idea of patients having access to their information, they expressed concerns around patients' understanding of the information contained in the medical record, especially access to laboratory or diagnostic test results [50,51,57,64,121,125]. One study reported an increase in the volume of inquiries and appointments due to patients not understanding the information. Another study found that providers had to spend more time reassuring patients after they read their records and medical notes [54]. A different study found that the majority of both physicians and nurses believed medical notes were confusing for patients, which had led to worry and increased contact between providers and patients [57]. With the intent to decrease patient confusion, few studies reflected on the aspect that portals had shifted how charting was done within the settings that offered access [42,54,66].

Many studies concluded that the health care workforce had a direct impact on portal adoption and utilization by patients [55,66,82,126]. Although paper access to medical records had always been obtainable and not seen as contentious, real-time access had been associated with concerns about privacy and security and led to discontent and low uptake by providers [120,123,124].

Only one study documented increased levels of threats and violence reported by staff from patients with access to their health records [40]. Staff reported that this was due to disagreement with the information in the medical record; however, the authors did not find increased incidents [40].

Impact on Workflow

Another concern was the workflow impact due to portal implementation. In one study, participating providers indicated that the portal implementation did not have a negative impact on their salary; however, 43% of the same participants believed that the portal increased their workload [41]. Similar findings were presented by Cajander et al [42], as the nurses in the study described how patients called and sent messages for the same question, which led to increased workload because of duplication of services. Furthermore, patients came prepared with more informed questions which led to *more in-depth discussions* that took additional unplanned appointment time [42].

Another study captured the experiences of providers who described cases in which patients contacted them for abnormalities that were clinically insignificant, thereby increasing the workload [125]. Vydra et al [66] compared provider reported time dedicated to portal-related duties with the administrative data captured by the system, as providers in their study reported spending an average of 12.5 hours per week logged into the portal; however, institutional records indicated an average of 8.2 hours per week.

Impact of Bidirectional Secure Messaging

Several studies captured the apprehensions of providers due to secure messaging [122,124]. In these studies, health care providers articulated the lack of clarity around appropriate ways to communicate via a portal as patients had expectations to receive immediate responses to their inquires [124]. Lieu et al [122] reported provider anxiety because of the lack of volume restrictions on electronic messages and their coping strategies to timing their responses to patient messages. Another study found that messaging increased work outside normal work hours [114].

A Canadian study found that providers considered to be early adopters indicated that secure messaging improved communication and interactions between themselves and their patients [114].

Virtual Care

One study reported that patient portals were effective for postoperative care, as patients uploaded images instead of scheduling face-to-face visits [87]. In this study, surgeons reported that web-based and clinic visits were equally effective for 68% (34/50) of patients.

Health System Perspective (Reduced Per Capita Cost of Health Care)

The health system perspective was explored by 15 peer-reviewed [45,48,49,59,65,69-71,128-134] and 3 gray literature [106,109,135] studies, which focused on the impact of patient portals on the potential for reducing costs. The studies varied in data collection approaches (prospective observational, n=4; retrospective observational, n=9; survey methods, n=2; and mixed methods, n=3), and the evaluated portal components (portal in general, n=11; viewing laboratory results, n=1; OpenNotes, n=1; secure messaging, n=4; appointments, n=2; and care plan, n=1). A total of 14 studies were from the United States, and there was 1 from Spain, Finland, and Canada. In all studies, EHR administrative data were used to compare pre- and postintervention inputs and outputs. In addition to the EHR, portal administrative data, workbench, interviews, charts, and tools that measured patient activation, quality of life, self-efficacy, and experience were applied. The studies explored whether implementation and subsequent adoption of a patient portal reduced costs in the evaluated areas. The included studies also explored whether patient portal use led to a decrease in 30-day all-cause readmission rates, clinical services utilization, no-show appointments, and the number of traditional encounters (office visits and telephone calls).

Decrease in No-Show Appointments

Studies evaluating the possible impact of patient portals on no-show appointments compared the pre-post no-show appointment rates and established that portal users had a lower rate of no-show appointments compared with nonusers; however, there was no difference in appointment cancellation rates [48,69,70,129,130]. Mendel et al [130] found portal enrollment increased once the clinic staff promoted the portal as a tool for appointment reminders, which was also associated with increased patient satisfaction [130]. However, once the portal started experiencing technical glitches, the decrease was maintained for only a short period of time. Zhong et al [69] found that no-show rates for portal users were 30% less than for nonusers and that frequent users of secure messaging and viewing laboratory test functions had the largest reduction in no-show rates. Similar findings were captured by Graham et al [48], who found a 53% relative reduction in no-show appointments in the clinics that piloted a patient portal, and the Ontario Shores Centre for Mental Health Sciences showed that portal users missed 18% of total appointments compared with nonportal users, who missed 20% of total appointments [135].

Impact on Office Visits and Telephone Calls

Several studies [128,132-134] evaluated the impact of secure messaging on office visits, telephone rates, and hospital readmissions. These studies had similar findings regardless of the applied methodology. Dexter et al [132] hypothesized that an increase in secure messaging use would decrease telephone call rates; however, the authors found that as messages increased, so did the number of telephone calls to the clinics. Similarly, Bryan et al [133] found that patients who sent more messages had higher clinical and phone encounters than those who did not send any. This impacted the workflow and the ability of care settings to handle the influx of visits and calls. Plate et al [128] demonstrated that if patients sent >2 messages and the clinic response rate was <75%, this significantly increased 90-day readmissions and 90-day emergency department visits.

Impact on Admission and Readmission Rates and Emergency Department Visits

Four studies [45,49,59,65] evaluated whether patient portals had the capacity to decrease readmission rates, emergency department visits, and hospital admissions. Nicolas et al [59] found a decrease in the rate of hospital admissions (5.28/10,000 per year) and 30-day all-cause readmissions (5.20/10,000 per year), whereas no increase in outpatient visit rates in the postintervention period of the portal implementation [59]. Similarly, Sorondo et al [65] concluded that there was a decrease in emergency department visits by at least 21% per 1000 and hospital admission rates by at least 38% per 1000. Although Nicolas et al [59] and Sorondo et al [65] found a decrease, Dumitrascu et al [45] and Griffin et al [49] concluded that portals users were associated with higher use of medical services and higher hospital readmissions.

Impact on Health Care Utilization

Four studies examined whether increased portal use decreased care utilization. Leveille et al [131] could not find any correlation between viewing medical notes and the number of appointments, whereas Zhou et al [71] concluded that viewing laboratory test results led to increased clinic visits and telephone calls. In their study, eHealth Saskatchewan reported that the number of primary care visits decreased because of patients using the portal [109]. This was also found in patient self-reported surveys. In their report, Canada Health Infoway reported an increase in e-visits because of patients having access to technologies providing e-service opportunities [106].

Mapping According to the BE Framework

Altogether, 77 peer-reviewed studies were mapped on the expanded version of the BE framework. The mapping was done by subdimension to create a more precise representation about the areas that are currently explored when studying patient portals. Most of the studies evaluated more than one subdimension. An overview of the number of studies per subdimension is shown in Multimedia Appendix 7. The numbers for each subdimension represent the number of studies that addressed a particular component.

The authors of 48 out of the 77 studies explored the ability of patient and providers to access services, followed by use behavior/pattern (n=48), user satisfaction (n=34), self-reported use (n=31), patient/caregiver participation (n=30), knowledge, attitude, perception, decision confidence, compliance (n=29), and ease of use (n=26).

The least measured changes related to security (n=1); data quality improvement, reduced loss/paper, and transcription errors (n=2); responsiveness (n=3), barriers, training, organizational support, time-to-evaluation, lessons, success factors, project management, leadership, and costs (n=4); functionality (n=5); and performance (n=6).

Detailed mapping of each study per BE framework subdimension is presented in Multimedia Appendix 7 [27,40-99,115-118,120-125,128-134].

Discussion

Principal Findings

The implementation of technologies such as patient portals is a convoluted process with ambiguous returns on investment. Frequently, implementation planning is intertwined with optimism related to *if we offer it, they will come*. However, the current realities show that the planning and implementation stages are based on very little preparation related to visualizing the identified need for the technology and the issues that it is trying to amend. On the basis of the results from this scoping review, several gaps in the literature were identified. For each documented gap, summary recommendations are provided on how to improve the measuring impact of patient portals based on the QA and BE frameworks.

Lack of Studies With Multidimensional Impact Evaluation Strategies

Although this study was based on comprehensive inclusion criteria, there were no peer-reviewed or non–peer-reviewed studies that measured impact based on all dimensions from the QA or BE frameworks. Although several reports describing the use of the BE framework were included, indicators and outcomes measured still focused on certain dimensions of the frameworks rather than most or all dimensions. From the 96 included studies, the most evaluated number of dimensions was 2. In the gray literature, however, studies often evaluated 3 of the 4 dimensions. When applying the same mapping process to included studies through the extended version of the BE framework [35], the greatest number of evaluated dimensions was 9 [123,125].

Multidimensional evaluation strategies require that research studies capture the patient experience (ie, patient satisfaction, patient engagement, convenience of care, care plan compliance, patient to care team ratio, and access to care), which is related to the population experience (ie, improved health outcomes, compliance with standards of care, insight about population health, and quality of life while reducing complications, mortality rate, hospital admissions, and emergency department visits). As patient and population experiences are interconnected with the health care provider experience (ie, reduced redundant tasks, burnout, and turnover rate while improving resource utilization, satisfaction, and provider-patient relationship), they need to be measured along with the health system experience (ie, reduced cancellations, staff costs, cost per patient, costs because of readmission and emergency department visits, length of hospital stay while developing improved opportunities for reimbursements). These outcomes also fit within the BE framework; however, measures of the impact in terms of system quality (ie, functionality, performance, and security), information quality (ie, content and availability), and service quality (ie, responsiveness) need to be developed and added.

Lack of Studies Based on Suitable Methodology and Sample Size to Evaluate Patient Portal Technology

A substantial gap in the literature was the lack of prospective longitudinal studies with large samples. There were a few prospective studies [55,65,85] that followed patients between 7 and 12 months; however, the sample sizes were small (between 20 and 94 participants). Nearly all studies that evaluated the patient perspective acknowledged that the study samples were made up of early adopters and individuals from the higher socioeconomic spectrum. Study limitations were the low study response and participation rates, the lack of ethically and racially diverse participants, and the lack of nonusers (patients and providers) perspectives.

Surveys used to measure satisfaction were *newly developed* or *developed based on previous studies*, and thus, there was little evidence of reliability and validity. The Canada Health Infoway System and Use Survey [136] was used by all BE studies; however, their reliability and validity were unclear.

In addition, during the review of the included studies, there were inconsistencies between terms such as *indicators*, *outcomes*, *measures*, *tools*, and *net benefits* as they relate to measuring patient portal impact. Although many studies used BE framework terminology, the concepts were applied in different ways.

Longitudinal studies of the impact of patient portals on patients would provide more real-world data about how users of portals interact and what potentially meaningful changes are needed. These types of studies could provide evidence about cause-and-effect relationships, which remain minimally explored from the standpoint of portal use and quality of care, satisfaction, communication, and health outcomes. Size and diversity in the patient and provider samples are key to envisaging solutions that would lead to use and eventually improve value-based care. In addition, using validated surveys would ensure that the right things are measured correctly. Measuring satisfaction with technology needs to encompass elements such as preference, proficiency, and performance.

Lack of Recognition and Evidence Utilization From Organizational and Health System–Level Internal Impact Evaluations of Patient Portals

During the gray literature search, many organizational and health system patient portal evaluation reports were identified. However, when efforts were made to find the corresponding peer-reviewed publications, none were found. As these BE reports (all from Canadian jurisdictions) measured the impact in several BE framework dimensions, it would be helpful for these studies and their findings to be acquired through peer-reviewed journals. Such publications would provide evidence on how to evaluate patient portal impact and would be shared more extensively. Furthermore, real-world impact evaluations would guide investigators in directing research that is deemed important by organizations and systems that implement health information technologies.

Lack of Operative Recommendations Based on Study Findings

Frequently, findings were not followed up by concrete recommendations as to what was needed to rectify the documented obstacles. For example, patient and provider satisfaction were considered important outcomes and measured through interviews or surveys; however, by stating that patients reported high satisfaction or providers reported low satisfaction, the studies did not elaborate on what the satisfaction levels meant in terms of changes or modifications. Are measures such as medium-to-high or high satisfaction enough to conclude that the evaluated patient portal was effective and should be maintained? Through this scoping review, the evidence showed that suggestions for change were based on technical or user change (ie, accessibility or increased provider use). The current evidence warrants recommendations for changes that can be effectively implemented and evaluated but require system change.

Lack of Use of Patient Self-Reported Health Outcomes

Only one study [65] reported using a patient self-reported health outcome tool (EuroQol-visual analogue scales [EQ-VAS]). Despite studies [43,44,46,53,96] showing correlations between chronic conditions, medication adherence, and use of patient portals, there is a lack of application and use of patient self-reported health outcome tools in patient portal studies that measure impact [137]. Patient self-reported outcomes have the capacity to improve the quality of patient care; however, they are very difficult to measure or capture through interviews. These types of outcome tools are cost-effective and can easily be integrated within the patient’s portal structure. As patient portals are implemented in diverse settings, the use of these types of tools would provide the ability to determine the unique threshold levels and plan for patient portal intervention strategies that would be more effective and appropriate for each setting. Measuring patient portal impact by incorporating patient self-reported health outcome tools would allow for condition-specific portal enhancements with the possibility of increased adoption and use.

Limitations

This scoping review included some literature that was not peer reviewed, and the strength of the evidence in these studies was not evaluated. First, the authors did not contact any research experts to help identify other gray literature. Second, only English language articles were included, which limited the databases and search terms used. For this reason, although a large number of citations were included, some relevant articles may have been missed. Third, as customary, scoping reviews do not quantitatively synthesize the findings; therefore, statistical conclusions may be drawn from the results regarding effects, statistical significance, or bias evaluation. Finally, study screening and selection are subjective processes. Although a high level of agreement was achieved, there was a reliance on judgment to include and exclude studies.

Conclusions

Despite extensive and existing research in the area of patient portals, the evidence from this scoping review suggests that impact research is available; however, it lacks multidimensionality. The QA and BE frameworks provided guidance in identifying the gaps in the current literature by providing a way to show how impact was assessed. This study highlights the need to appropriately plan how impact will be assessed and how the findings will be translated into effective adaptations. If the *how* and *what* are not properly planned, the generalizability of patient portal studies will continue to elude researchers and implementation teams.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Scoping review search strategy.

[DOCX File , 19 KB - jmir_v22i12e24568_app1.docx]

Multimedia Appendix 2

Synthesis of included studies.

[DOCX File , 25 KB - jmir_v22i12e24568_app2.docx]

Multimedia Appendix 3

Quadruple Aim patient perspective table.

[DOCX File , 44 KB - jmir_v22i12e24568_app3.docx]

Multimedia Appendix 4

Quadruple Aim population perspective table.

[DOCX File , 21 KB - jmir_v22i12e24568_app4.docx]

Multimedia Appendix 5

Quadruple Aim healthcare provider perspective table.

[DOCX File , 25 KB - jmir_v22i12e24568_app5.docx]

Multimedia Appendix 6
Quadruple Aim Health system perspective table.
[DOCX File , 22 KB - jmir_v22i12e24568_app6.docx]

Multimedia Appendix 7
Mapping the peer-reviewed studies according to the Benefits Evaluation Framework.
[DOCX File , 35 KB - jmir_v22i12e24568_app7.docx]

References

1. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
2. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769. [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
3. Ancker JS, Witteman HO, Hafeez B, Provencher T, van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. J Med Internet Res 2015 Jun 04;17(6):e137 [FREE Full text] [doi: 10.2196/jmir.4381] [Medline: 26043709]
4. Kerns JW, Krist AH, Longo DR, Kuzel AJ, Woolf SH. How patients want to engage with their personal health record: a qualitative study. BMJ Open 2013 Jul 30;3(7):- [FREE Full text] [doi: 10.1136/bmjopen-2013-002931] [Medline: 23901027]
5. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
6. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015 Dec 3;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
7. Zhou L, Soran CS, Jenter CA, Volk LA, Orav EJ, Bates DW, et al. The relationship between electronic health record use and quality of care over time. J Am Med Inform Assoc 2009;16(4):457-464 [FREE Full text] [doi: 10.1197/jamia.M3128] [Medline: 19390094]
8. Patel V, Jamoom E, Hsiao C, Furukawa MF, Buntin M. Variation in electronic health record adoption and readiness for meaningful use: 2008-2011. J Gen Intern Med 2013 Jul;28(7):957-964 [FREE Full text] [doi: 10.1007/s11606-012-2324-x] [Medline: 23371416]
9. Bourgeois FC, Taylor PL, Emans SJ, Nigrin DJ, Mandl KD. Whose personal control? Creating private, personally controlled health records for pediatric and adolescent patients. J Am Med Inform Assoc 2008;15(6):737-743 [FREE Full text] [doi: 10.1197/jamia.M2865] [Medline: 18755989]
10. Xierali IM, Hsiao C, Puffer JC, Green LA, Rinaldo JC, Bazemore AW, et al. The rise of electronic health record adoption among family physicians. Ann Fam Med 2013;11(1):14-19 [FREE Full text] [doi: 10.1370/afm.1461] [Medline: 23319501]
11. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
12. Fuji KT, Abbott AA, Galt KA. Personal health record design: qualitative exploration of issues inhibiting optimal use. Diabetes Care 2014;37(1):e13-e14 [FREE Full text] [doi: 10.2337/dc13-1630] [Medline: 24356602]
13. Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
14. Webster PC. Canada's electronic health records initiative stalled by federal funding freeze. Can Med Assoc J 2010 Mar 23;182(5):E247-E248 [FREE Full text] [doi: 10.1503/cmaj.109-3183] [Medline: 20159901]
15. Witry MJ, Doucette WR, Daly JM, Levy BT, Chrischilles EA. Family physician perceptions of personal health records. Perspect Health Inf Manag 2010 Jan 1;7:1d [FREE Full text] [Medline: 20697465]
16. Webster PC. National standards for electronic health records remain remote. Can Med Assoc J 2010 Jun 15;182(9):888-889 [FREE Full text] [doi: 10.1503/cmaj.109-3239] [Medline: 20403893]
17. Tu K, Widdifield J, Young J, Oud W, Ivers NM, Butt DA, et al. Are family physicians comprehensively using electronic medical records such that the data can be used for secondary purposes? A Canadian perspective. BMC Med Inform Decis Mak 2015 Aug 13;15:67 [FREE Full text] [doi: 10.1186/s12911-015-0195-x] [Medline: 26268511]
18. Yau GL, Williams AS, Brown JB. Family physicians' perspectives on personal health records: qualitative study. Can Fam Physician 2011 May;57(5):e178-e184 [FREE Full text] [Medline: 21642732]
19. Webster PC. The pocketbook impact of electronic health records: part 1. Fee-for-service billing is compatible with EHRs: agreed. Can Med Assoc J 2010 May 18;182(8):752-753 [FREE Full text] [doi: 10.1503/cmaj.109-3225] [Medline: 20371650]

20. Avdagovska M, Stafinski T, Ballermann M, Menon D, Olson K, Paul P. Tracing the decisions that shaped the development of MyChart, an electronic patient portal in Alberta, Canada: historical research study. J Med Internet Res 2020 May 26;22(5):e17505 [FREE Full text] [doi: 10.2196/17505] [Medline: 32452811]
21. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
22. Bates DW. The quality case for information technology in healthcare. BMC Med Inform Decis Mak 2002 Oct 23;2:7 [FREE Full text] [doi: 10.1186/1472-6947-2-7] [Medline: 12396233]
23. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
24. Han H, Gleason KT, Sun C, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
25. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
26. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
27. Zanaboni P, Ngangue P, Mbemba GI, Schopf TR, Bergmo TS, Gagnon M. Methods to evaluate the effects of internet-based digital health interventions for citizens: systematic review of reviews. J Med Internet Res 2018 Jun 7;20(6):e10202 [FREE Full text] [doi: 10.2196/10202] [Medline: 29880470]
28. Lester M, Boateng S, Studeny J, Coustasse A. Personal health records: beneficial or burdensome for patients and healthcare providers? Perspect Health Inf Manag 2016;13:1h [FREE Full text] [Medline: 27134613]
29. Thompson MJ, Reilly JD, Valdez RS. Work system barriers to patient, provider, and caregiver use of personal health records: a systematic review. Appl Ergon 2016 May;54:218-242. [doi: 10.1016/j.apergo.2015.10.010] [Medline: 26851482]
30. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient portal use and experience among older adults: systematic review. JMIR Med Inform 2017 Oct 16;5(4):e38 [FREE Full text] [doi: 10.2196/medinform.8092] [Medline: 29038093]
31. Jilka SR, Callahan R, Sevdalis N, Mayer EK, Darzi A. 'Nothing about me without me': an interpretative review of patient accessible electronic health records. J Med Internet Res 2015 Jun 29;17(6):e161 [FREE Full text] [doi: 10.2196/jmir.4446] [Medline: 26123476]
32. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
33. Mold F, Raleigh M, Alharbi NS, de Lusignan S. The impact of patient online access to computerized medical records and services on type 2 diabetes: systematic review. J Med Internet Res 2018 Jul 6;20(7):e235 [FREE Full text] [doi: 10.2196/jmir.7858] [Medline: 29980499]
34. Lau F, Hagens S, Muttitt S. A proposed benefits evaluation framework for health information systems in Canada. Healthc Q 2007;10(1):112-6, 118. [Medline: 17326376]
35. Lau F, Kuziemsky C, Price M, Gardner J. A review on systematic reviews of health information system studies. J Am Med Inform Assoc 2010;17(6):637-645 [FREE Full text] [doi: 10.1136/jamia.2010.004838] [Medline: 20962125]
36. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Method 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
37. Health Impact Assessement. Health Education Partners. 1999. URL: http://www.healthedpartners.org/ceu/hia/hia01/01_02_gothenburg_paper_on_hia_1999.pdf [accessed 2020-05-20]
38. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. Press peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
39. Anderson S, Allen P, Peckham S, Goodwin N. Asking the right questions: scoping studies in the commissioning of research on the organisation and delivery of health services. Health Res Policy Syst 2008 Jul 9;6:7 [FREE Full text] [doi: 10.1186/1478-4505-6-7] [Medline: 18613961]
40. Åkerstedt US, Cajander A, Moll J, Ålander T. On threats and violence for staff and patient accessible electronic health records. Cogent Psychol 2018 Sep 6;5(1):1-11. [doi: 10.1080/23311908.2018.1518967]
41. Bush RA, Connelly CD, Pérez A, Chan N, Kuelbs C, Chiang GJ. Physician perception of the role of the patient portal in pediatric health. J Ambul Care Manage 2017;40(3):238-245 [FREE Full text] [doi: 10.1097/JAC.0000000000000175] [Medline: 28350637]
42. Cajander A, Moll J, Englund S, Hansman A. Medical records online for patients and effects on the work environment of nurses. Stud Health Technol Inform 2018;247:271-275. [Medline: 29677965]

43. Crouch PB, Rose CD, Johnson M, Janson SL. A pilot study to evaluate the magnitude of association of the use of electronic personal health records with patient activation and empowerment in HIV-infected veterans. PeerJ 2015;3:e852 [FREE Full text] [doi: 10.7717/peerj.852] [Medline: 25802815]
44. Devkota B, Salas J, Sayavong S, Scherrer JF. Use of an online patient portal and glucose control in primary care patients with diabetes. Popul Health Manag 2016 Apr;19(2):125-131. [doi: 10.1089/pop.2015.0034] [Medline: 26237200]
45. Dumitrascu A, Burton M, Dawson N, Thomas C, Nordan L, Greig H, et al. Patient portal use and hospital outcomes. J Am Med Inform Assoc 2018 Apr 1;25(4):447-453. [doi: 10.1093/jamia/ocx149] [Medline: 29300961]
46. Fiks AG, Mayne SL, Karavite DJ, Suh A, O'Hara R, Localio AR, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973 [FREE Full text] [doi: 10.1542/peds.2014-3167] [Medline: 25755233]
47. Megan Forster M, Dennison K, Callen J, Andrew A, Westbrook JI. Maternity patients' access to their electronic medical records: use and perspectives of a patient portal. Health Inf Manag 2015;44(1):4-11. [doi: 10.1177/183335831504400101] [Medline: 27092464]
48. Graham TA, Ali S, Avdagovska M, Ballermann M. Effects of a web-based patient portal on patient satisfaction and missed appointment rates: survey study. J Med Internet Res 2020 May 19;22(5):e17955 [FREE Full text] [doi: 10.2196/17955] [Medline: 32427109]
49. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient portals: who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2016;7(2):489-501 [FREE Full text] [doi: 10.4338/ACI-2016-01-RA-0003] [Medline: 27437056]
50. Grossman LV, Masterson Creber RM, Ryan B, Restaino S, Alarcon I, Polubriaginof F, et al. Providers' perspectives on sharing health information through acute care patient portals. AMIA Annu Symp Proc 2018;2018:1273-1281 [FREE Full text] [Medline: 30815169]
51. Grünloh C, Myreteg G, Cajander A, Rexhepi H. 'Why do they need to check me?' patient participation through ehealth and the doctor-patient relationship: qualitative study. J Med Internet Res 2018 Jan 15;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]
52. Hanna L, Gill SD, Newstead L, Hawkins M, Osborne RH. Patient perspectives on a personally controlled electronic health record used in regional Australia. Health Inf Manag 2017 Jan;46(1):42-48. [doi: 10.1177/1833358316661063] [Medline: 27486184]
53. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in electronic health record patient portal use in nephrology clinics. Clin J Am Soc Nephrol 2015 Nov 6;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]
54. Johansen MA, Kummervold PE, Sørensen T, Zanaboni P. Health professionals' experience with patients accessing their electronic health records: results from an online survey. Stud Health Technol Inform 2019 Aug 21;264:504-508. [doi: 10.3233/SHTI190273] [Medline: 31437974]
55. King G, Maxwell J, Karmali A, Hagens S, Pinto M, Williams L, et al. Connecting families to their health record and care team: the use, utility, and impact of a client/family health portal at a children's rehabilitation hospital. J Med Internet Res 2017 Apr 6;19(4):e97 [FREE Full text] [doi: 10.2196/jmir.6811] [Medline: 28385680]
56. Manard W, Scherrer JF, Salas J, Schneider FD. Patient portal use and blood pressure control in newly diagnosed hypertension. J Am Board Fam Med 2016;29(4):452-459 [FREE Full text] [doi: 10.3122/jabfm.2016.04.160008] [Medline: 27390376]
57. Moll J, Cajander A. Oncology health-care professionals' perceived effects of patient accessible electronic health records 6 years after launch: A survey study at a major university hospital in Sweden. Health Informatics J 2020 Jun;26(2):1392-1403 [FREE Full text] [doi: 10.1177/1460458219881007] [Medline: 31621459]
58. Moll J, Rexhepi H, Cajander A, Grünloh C, Huvila I, Hägglund M, et al. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. J Med Internet Res 2018 Nov 1;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
59. Martínez Nicolás I, lê Cook B, Flores M, del Olmo Rodriguez M, Hernández Rodríguez C, Llamas Sillero P, et al. The impact of a comprehensive electronic patient portal on the health service use: an interrupted time-series analysis. Eur J Public Health 2019 Jun 1;29(3):413-418. [doi: 10.1093/eurpub/cky257] [Medline: 30544169]
60. Reed ME, Huang J, Millman A, Graetz I, Hsu J, Brand R, et al. Portal use among patients with chronic conditions: patient-reported care experiences. Med Care 2019 Oct;57(10):809-814. [doi: 10.1097/MLR.0000000000001178] [Medline: 31415340]
61. Rief JJ, Hamm ME, Zickmund SL, Nikolajski C, Lesky D, Hess R, et al. Using health information technology to foster engagement: patients' experiences with an active patient health record. Health Commun 2017 Mar;32(3):310-319. [doi: 10.1080/10410236.2016.1138378] [Medline: 27223684]
62. Ronda MC, Dijkhorst-Oei L, Rutten GE. Patients' experiences with and attitudes towards a diabetes patient web portal. PLoS One 2015;10(6):e0129403 [FREE Full text] [doi: 10.1371/journal.pone.0129403] [Medline: 26086272]
63. Shah SG, Fitton R, Hannan A, Fisher B, Young T, Barnett J. Accessing personal medical records online: a means to what ends? Int J Med Inform 2015 Feb;84(2):111-118 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.005] [Medline: 25453275]

64. Shaw CL, Casterline GL, Taylor D, Fogle M, Granger B. Increasing health portal utilization in cardiac ambulatory patients: a pilot project. Comput Inform Nurs 2017 Oct;35(10):512-519. [doi: 10.1097/CIN.0000000000000361] [Medline: 28541965]
65. Sorondo B, Allen A, Fathima S, Bayleran J, Sabbagh I. Patient portal as a tool for enhancing patient experience and improving quality of care in primary care practices. EGEMS (Wash DC) 2016;4(1):1262 [FREE Full text] [doi: 10.13063/2327-9214.1262] [Medline: 28203611]
66. Vydra TP, Cuaresma E, Kretovics M, Bose-Brill S. Diffusion and use of tethered personal health records in primary care. Perspect Health Inf Manag 2015;12:1c [FREE Full text] [Medline: 26755897]
67. Wildenbos GA, Maasri K, Jaspers M, Peute L. Older adults using a patient portal: registration and experiences, one year after implementation. Digit Health 2018;4:2055207618797883 [FREE Full text] [doi: 10.1177/2055207618797883] [Medline: 30186619]
68. Zanaboni P, Kummervold PE, Sørensen T, Johansen MA. Patient use and experience with online access to electronic health records in Norway: results from an online survey. J Med Internet Res 2020 Feb 7;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]
69. Zhong X, Park J, Liang M, Shi F, Budd PR, Sprague JL, et al. Characteristics of patients using different patient portal functions and the impact on primary care service utilization and appointment adherence: retrospective observational study. J Med Internet Res 2020 Feb 25;22(2):e14410 [FREE Full text] [doi: 10.2196/14410] [Medline: 32130124]
70. Zhong X, Liang M, Sanchez R, Yu M, Budd PR, Sprague JL, et al. On the effect of electronic patient portal on primary care utilization and appointment adherence. BMC Med Inform Decis Mak 2018 Oct 16;18(1):84 [FREE Full text] [doi: 10.1186/s12911-018-0669-8] [Medline: 30326876]
71. Zhou YY, Leith WM, Li H, Tom JO. Personal health record use for children and health care utilization: propensity score-matched cohort analysis. J Am Med Inform Assoc 2015 Jul;22(4):748-754. [doi: 10.1093/jamia/ocu018] [Medline: 25656517]
72. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What patients value about reading visit notes: a qualitative inquiry of patient experiences with their health information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]
73. Giardina TD, Baldwin J, Nystrom DT, Sittig DF, Singh H. Patient perceptions of receiving test results via online portals: a mixed-methods study. J Am Med Inform Assoc 2018 Apr 1;25(4):440-446 [FREE Full text] [doi: 10.1093/jamia/ocx140] [Medline: 29240899]
74. Kayastha N, Pollak KI, LeBlanc TW. Open oncology notes: a qualitative study of oncology patients' experiences reading their cancer care notes. J Oncol Pract 2018 Apr;14(4):e251-e258. [doi: 10.1200/jop.2017.028605]
75. Denneson LM, Chen JI, Pisciotta M, Tuepker A, Dobscha SK. Patients' positive and negative responses to reading mental health clinical notes online. Psychiatr Serv 2018 May 1;69(5):593-596. [doi: 10.1176/appi.ps.201700353] [Medline: 29493408]
76. Chimowitz H, Gerard M, Fossa A, Bourgeois F, Bell SK. Empowering informal caregivers with health information: opennotes as a safety strategy. Jt Comm J Qual Patient Saf 2018 Mar;44(3):130-136. [doi: 10.1016/j.jcjq.2017.09.004] [Medline: 29499809]
77. Robinson S, Reed M, Quevillon T, Hirvi E. Patient perceptions and interactions with their web portal-based laboratory results. BMJ Health Care Inform 2019 Apr;26(1) [FREE Full text] [doi: 10.1136/bmjhci-2019-000012] [Medline: 31039117]
78. Fossa AJ, Bell SK, DesRoches C. OpenNotes and shared decision making: a growing practice in clinical transparency and how it can support patient-centered care. J Am Med Inform Assoc 2018 Sep 1;25(9):1153-1159. [doi: 10.1093/jamia/ocy083] [Medline: 29982659]
79. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. Opennotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 6;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
80. Dalal AK, Dykes P, Samal L, McNally K, Mlaver E, Yoon CS, et al. Potential of an electronic health record-integrated patient portal for improving care plan concordance during acute care. Appl Clin Inform 2019 May;10(3):358-366 [FREE Full text] [doi: 10.1055/s-0039-1688831] [Medline: 31141830]
81. Graetz I, Huang J, Muelly ER, Fireman B, Hsu J, Reed ME. Association of mobile patient portal access with diabetes medication adherence and glycemic levels among adults with diabetes. JAMA Netw Open 2020 Feb 5;3(2):e1921429 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.21429] [Medline: 32074289]
82. Wang J, Huang J, Cheung CS, Wong WN, Cheung NT, Wong MC. Adoption of an electronic patient record sharing pilot project: cross-sectional survey. J Med Internet Res 2020 Apr 6;22(4):e13761 [FREE Full text] [doi: 10.2196/13761] [Medline: 32250279]
83. Garry K, Blecker S, Saag H, Szerencsy A, Jones SA, Testa P, et al. Patient experience with notification of radiology results: a comparison of direct communication and patient portal use. J Am Coll Radiol 2020 Sep;17(9):1130-1138. [doi: 10.1016/j.jacr.2020.01.046] [Medline: 32289281]
84. Crotty BH, Mostaghimi A, O'Brien J, Bajracharya A, Safran C, Landon BE. Prevalence and risk profile of unread messages to patients in a patient web portal. Appl Clin Inform 2015;6(2):375-382 [FREE Full text] [doi: 10.4338/ACI-2015-01-CR-0006] [Medline: 26171082]

85. Saberi P, Catz SL, Leyden WA, Stewart C, Ralston JD, Horberg MA, et al. Antiretroviral therapy adherence and use of an electronic shared medical record among people living with HIV. AIDS Behav 2015 Jun;19(Suppl 2):177-185 [FREE Full text] [doi: 10.1007/s10461-014-0982-x] [Medline: 25572829]
86. Wright E, Darer J, Tang X, Thompson J, Tusing L, Fossa A, et al. Sharing physician notes through an electronic portal is associated with improved medication adherence: quasi-experimental study. J Med Internet Res 2015 Oct 8;17(10):e226 [FREE Full text] [doi: 10.2196/jmir.4872] [Medline: 26449757]
87. Broman KK, Oyefule OO, Phillips SE, Baucom RB, Holzman MD, Sharp KW, et al. Postoperative care using a secure online patient portal: changing the (inter)face of general surgery. J Am Coll Surg 2015 Dec;221(6):1057-1066 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2015.08.429] [Medline: 26453260]
88. Haun JN, Patel NR, Lind JD, Antinori N. Large-scale survey findings inform patients' experiences in using secure messaging to engage in patient-provider communication and self-care management: a quantitative assessment. J Med Internet Res 2015 Dec 21;17(12):e282 [FREE Full text] [doi: 10.2196/jmir.5152] [Medline: 26690761]
89. Raghu TS, Frey K, Chang Y, Cheng M, Freimund S, Patel A. Using secure messaging to update medications list in ambulatory care setting. Int J Med Inform 2015 Oct;84(10):754-762. [doi: 10.1016/j.ijmedinf.2015.06.003] [Medline: 26113460]
90. Giardina TD, Modi V, Parrish DE, Singh H. The patient portal and abnormal test results: an exploratory study of patient experiences. Patient Exp J 2015;2(1):148-154 [FREE Full text] [Medline: 28345018]
91. Reed M, Graetz I, Gordon N, Fung V. Patient-initiated e-mails to providers: associations with out-of-pocket visit costs, and impact on care-seeking and health. Am J Manag Care 2015 Dec 1;21(12):e632-e639 [FREE Full text] [Medline: 26760425]
92. Esch T, Mejilla R, Anselmo M, Podtschaske B, Delbanco T, Walker J. Engaging patients through open notes: an evaluation using mixed methods. BMJ Open 2016 Jan 29;6(1):e010034 [FREE Full text] [doi: 10.1136/bmjopen-2015-010034] [Medline: 26826154]
93. Mafi JN, Mejilla R, Feldman H, Ngo L, Delbanco T, Darer J, et al. Patients learning to read their doctors' notes: the importance of reminders. J Am Med Inform Assoc 2016 Sep;23(5):951-955 [FREE Full text] [doi: 10.1093/jamia/ocv167] [Medline: 26911830]
94. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained use of patient portal features and improvements in diabetes physiological measures. J Med Internet Res 2016 Jul 1;18(7):e179 [FREE Full text] [doi: 10.2196/jmir.5663] [Medline: 27369696]
95. Petullo B, Noble B, Dungan KM. Effect of electronic messaging on glucose control and hospital admissions among patients with diabetes. Diabetes Technol Ther 2016 Sep;18(9):555-560 [FREE Full text] [doi: 10.1089/dia.2016.0105] [Medline: 27398824]
96. Millman MD, Den Hartog KS. Optimizing adherence through provider and patient messaging. Popul Health Manag 2016 Aug;19(4):264-271. [doi: 10.1089/pop.2015.0068] [Medline: 26451807]
97. Reicher JJ, Reicher MA. Implementation of certified EHR, patient portal, and 'direct' messaging technology in a radiology environment enhances communication of radiology results to both referring physicians and patients. J Digit Imaging 2016 Jun;29(3):337-340 [FREE Full text] [doi: 10.1007/s10278-015-9845-x] [Medline: 26588906]
98. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2017 Apr 1;24(e1):e166-e172. [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
99. Peremislov D. Patient use of the electronic communication portal in management of type 2 diabetes. Comput Inform Nurs 2017 Sep;35(9):473-482. [doi: 10.1097/CIN.0000000000000348] [Medline: 28323648]
100. Masterson CR, Courtney LR, Ancker J, Dymek C. A National Web Conference on Effective Design and Use of Patient Portals and Their Impact on Patient-Centered Care. Agency for Healthcare Research and Quality. 2017. URL: https://digital.ahrq.gov/sites/default/files/docs/impact-pcc-slides-032317.pdf [accessed 2020-03-12]
101. Access to Digital Health Services 2019 Survey of Canadians Summary Report. Canada Health Infoway. 2019. URL: https://www.infoway-inforoute.ca/en/component/edocman/3786-access-to-digital-health-services-2019-survey-of-canadians-summary-report/view-document?Itemid=101 [accessed 2020-03-12]
102. A New Day in Health Care is Coming: Annual Report 2018-2019. Canada Health Infoway. 2019. URL: https://www.infoway-inforoute.ca/en/component/edocman/3726-annual-report-2018-2019/view-document?Itemid=101 [accessed 2020-03-10]
103. Impacts of Direct Patient Access to Laboratory Results: Final Report. The Social Research and Demonstration Corporation. 2014. URL: https://www.infoway-inforoute.ca/en/component/edocman/2775-impacts-of-direct-patient-access-to-laboratory-results-final-report/view-document?Itemid=0 [accessed 2020-03-12]
104. Dyke E. Environmental Scan: Processes to Enable Adolescent Access to Personal Health Records. Canada Health Infoway. 2016. URL: https://www.researchgate.net/publication/309232044 [accessed 2020-03-12]
105. Defining Empowerment and Supporting Engagement Saskatchewan Patients and the Ehealth Saskatchewan Citizen Health Information Portal Chip Internet. eHealth Saskatchewan. 2016. URL: https://www.infoway-inforoute.ca/en/component/edocman/

3406-defining-empowerment-and-supporting-engagement-saskatchewan-patients-and-the-ehealth-saskatchewan-citizen-health-information-portal-chip/view-document?Itemid=101 [accessed 2020-03-10]
106. Backgrounder - Positive Patient Experience Yields Health Care Benefits. Canada Health Information. 2016. URL: https://www.infoway-inforoute.ca/en/component/edocman/3254-backgrounder-positive-patient-experience-yields-health-care-benefits/view-document?Itemid=101 [accessed 2020-03-10]
107. Valuing Canadians' Secure Access to Their Health Information and Digital Health eServices. Canada Health Infoway. 2018. URL: https://www.infoway-inforoute.ca/en/component/edocman/3552-valuing-canadians-secure-access-to-their-health-information-and-digital-health-eservices/view-document?Itemid=101 [accessed 2020-03-10]
108. Canada Health Infoway: Digital Health in Canada. 2018. URL: https://www.infoway-inforoute.ca/en/component/edocman/3542-community-paramedicine-remote-patient-monitoring-cprpm-benefits-evaluation-lessons-learned/view-document?Itemid=10 [accessed 2020-03-10]
109. Citizen Health Information Portal Chip: Benefits Evaluation Report V1. Canada Health Infoway. 2017. URL: https://www.infoway-inforoute.ca/en/component/edocman/3402-ehealth-saskatchewan-citizen-health-information-portal-chip-benefits-evaluation-report/view-document?Itemid=101 [accessed 2020-03-10]
110. Connect2Care Benefits Evaluation-Results and Final Report. Holland Bloorview. 2016. URL: https://www.infoway-inforoute.ca/en/component/edocman/3170-holland-bloorview-connect2care-benefits-evaluation-report/view-document?Itemid=101 [accessed 2020-03-10]
111. Epic EHR Program: MyChart Consumer Health Solutions Benefits Evaluation Report Pilot CHS-09B / CHSONCHEO02. Children's Hospital of Eastern Ontario. 2015. URL: https://www.infoway-inforoute.ca/en/component/edocman/3091-epic-ehr-program-mychart-consumer-health-solutions-benefits-evaluation-report-pilot/view-document?Itemid=101 [accessed 2020-03-10]
112. miDASH , Consumer Health Solution Canada Health Infoway Benefits Evaluation. Health Quality Innovation Collaborative. 2016. URL: https://www.infoway-inforoute.ca/en/component/edocman/3137-midash-consumer-health-solution-benefits-evaluation/view-document?Itemid=101 [accessed 2020-03-10]
113. myCARE Benefits Evaluation and Final Report. Group Health Centre. 2016. URL: https://www.infoway-inforoute.ca/en/component/edocman/3129-mycare-benefits-evaluation-and-final-report/view-document?Itemid=101 [accessed 2020-03-10]
114. Project Review and Benefits Evaluation Final Report Consumer Health Solution for Patient and Provider Communication. Barrie Community Health Link. 2016. URL: https://www.infoway-inforoute.ca/en/component/edocman/3150-myhealthlinked-project-review-and-benefits-evaluation-final-report/view-document?Itemid=101 [accessed 2020-03-10]
115. Cutrona SL, Golden JG, Goff SL, Ogarek J, Barton B, Fisher L, et al. Improving rates of outpatient influenza vaccination through EHR portal messages and interactive automated calls: a randomized controlled trial. J Gen Intern Med 2018 May;33(5):659-667 [FREE Full text] [doi: 10.1007/s11606-017-4266-9] [Medline: 29383550]
116. Lyles CR, Sarkar U, Schillinger D, Ralston JD, Allen JY, Nguyen R, et al. Refilling medications through an online patient portal: consistent improvements in adherence across racial/ethnic groups. J Am Med Inform Assoc 2016 Apr;23(e1):e28-e33 [FREE Full text] [doi: 10.1093/jamia/ocv126] [Medline: 26335983]
117. Szilagyi P, Albertin C, Casillas A, Valderrama R, Duru O, Ong M, et al. Effect of patient portal reminders sent by a health care system on influenza vaccination rates: a randomized clinical trial. JAMA Intern Med 2020 Jul 1;180(7):962-970. [doi: 10.1001/jamainternmed.2020.1602] [Medline: 32421168]
118. Foster B, Krasowski M. The use of an electronic health record patient portal to access diagnostic test results by emergency patients at an academic medical center: Retrospective study. J Med Internet Res 2019;216:1-14. [doi: 10.2196/preprints.13791]
119. Current and Potential Value of Canadians' Secure Access to Their Health Information and Digital Health Eservices Today's Presentation. Canada Health Infoway. 2018. URL: https://www.infoway-inforoute.ca/en/what-we-do/news-events/webinars/3548-current-and-potential-value-of-canadians-secure-access-to-their-health-information-and-digital-health-e-services/download [accessed 2020-03-10]
120. Grünloh C, Cajander A, Myreteg G. 'The record is our work tool!'-physicians' framing of a patient portal in Sweden. J Med Internet Res 2016 Jun 27;18(6):e167 [FREE Full text] [doi: 10.2196/jmir.5705] [Medline: 27349531]
121. Winget M, Haji-Sheikhi F, Brown-Johnson C, Rosenthal EL, Sharp C, Buyyounouski MK, et al. Electronic release of pathology and radiology results to patients: opinions and experiences of oncologists. J Oncol Pract 2016 Aug;12(8):e792-e799. [doi: 10.1200/jop.2016.011098]
122. Lieu TA, Altschuler A, Weiner JZ, East JA, Moeller MF, Prausnitz S, et al. Primary care physicians' experiences with and strategies for managing electronic messages. JAMA Netw Open 2019 Dec 2;2(12):e1918287 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18287] [Medline: 31880798]
123. Federman AD, Sanchez-Munoz A, Jandorf L, Salmon C, Wolf MS, Kannry J. Patient and clinician perspectives on the outpatient after-visit summary: a qualitative study to inform improvements in visit summary design. J Am Med Inform Assoc 2017 Apr 1;24(e1):e61-e68. [doi: 10.1093/jamia/ocw106] [Medline: 27497793]

124. Sieck CJ, Hefner JL, Schnierle J, Florian H, Agarwal A, Rundell K, et al. The rules of engagement: perspectives on secure messaging from experienced ambulatory patient portal users. JMIR Med Inform 2017 Jul 4;5(3):e13 [FREE Full text] [doi: 10.2196/medinform.7516] [Medline: 28676467]
125. Pillemer F, Price RA, Paone S, Martich GD, Albert S, Haidari L, et al. Direct release of test results to patients increases patient engagement and utilization of care. PLoS One 2016;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
126. aEPR Benefits Evaluation. Canada Health Infoway. 2016. URL: https://www.infoway-inforoute.ca/en/component/edocman/3248-women-s-college-hospital-ambulatory-electronic-patient-record-benefits-evaluation/view-document?Itemid=101 [accessed 2020-03-10]
127. Understanding the Current State of Patient- Provided Digital Health Information Know Me. Canada Health Infoway. 2015. URL: https://www.infoway-inforoute.ca/en/component/edocman/2880-understanding-the-current-state-of-patient-provided-digital-health-information-know-me/view-document?Itemid=101 [accessed 2020-03-10]
128. Plate JF, Ryan SP, Bergen MA, Hong CS, Attarian DE, Seyler TM. Utilization of an electronic patient portal following total joint arthroplasty does not decrease readmissions. J Arthroplasty 2019 Feb;34(2):211-214. [doi: 10.1016/j.arth.2018.11.002] [Medline: 30497899]
129. Shah K, Alshammaa A, Affan M, Schultz L, Walbert T, Zaman I. Education research: electronic patient portal enrollment and no-show rates within a neurology resident clinic. Neurology 2019 Jan 1;92(1):50-54. [doi: 10.1212/WNL.0000000000006685] [Medline: 30584078]
130. Mendel A, Chow S. Impact of health portal enrolment with email reminders at an academic rheumatology clinic. BMJ Qual Improv Rep 2017;6(1):- [FREE Full text] [doi: 10.1136/bmjquality.u214811.w5926] [Medline: 28321302]
131. Leveille SG, Mejilla R, Ngo L, Fossa A, Elmore JG, Darer J, et al. Do patients who access clinical information on patient internet portals have more primary care visits? Med Care 2016 Jan;54(1):17-23. [doi: 10.1097/MLR.0000000000000442] [Medline: 26565525]
132. Dexter EN, Fields S, Rdesinski RE, Sachdeva B, Yamashita D, Marino M. Patient-provider communication: does electronic messaging reduce incoming telephone calls? J Am Board Fam Med 2016;29(5):613-619 [FREE Full text] [doi: 10.3122/jabfm.2016.05.150371] [Medline: 27613794]
133. Bryan M, Norton D, Birstler J, Chen G, Cruz L, Hanrahan L. Resource utilization among portal users who send messages: a retrospective cohort study. Wis Med J 2020 Mar;119(1):26-32 [FREE Full text] [Medline: 32348068]
134. Riippa I, Linna M, Rönkkö I. A patient portal with electronic messaging: controlled before-and-after study. J Med Internet Res 2015 Nov 9;17(11):e250 [FREE Full text] [doi: 10.2196/jmir.4487] [Medline: 26553595]
135. HealthCheck Patient Portal, Ontario Shores Centre for Mental Health Sciences, Benefits Evaluation Report. Ontario Shores Centre for Mental Health Sciences. 2016. URL: https://www.infoway-inforoute.ca/en/component/edocman/3174-ontario-shores-healthcheck-patient-portal-benefits-evaluation-report/view-document?Itemid=101 [accessed 2020-04-09]
136. Canada Health Infoway. 2015. URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/toolkits/change-management/national-framework/monitoring-and-evaluation/resources-and-tools/991-benefits-evaluation-survey-process-system-use-assessment-survey [accessed 2020-02-05]
137. Anker S, Agewall S, Borggrefe M, Calvert M, Jaime Caro J, Cowie M, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. Eur Heart J 2014 Aug 7;35(30):2001-2009. [doi: 10.1093/eurheartj/ehu205] [Medline: 24904027]

Abbreviations

BE: Benefits Evaluation
CINAHL: Cumulative Index of Nursing and Allied Health Literature
EHR: electronic health record
HbA_{1c}: glycated hemoglobin
PRESS: peer review of electronic search strategies
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
QA: Quadruple Aim
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 25.09.20; peer-reviewed by D Juzwishin, R Fitton; comments to author 09.10.20; revised version received 15.10.20; accepted 28.10.20; published 08.12.20.

Please cite as:
Avdagovska M, Menon D, Stafinski T
Capturing the Impact of Patient Portals Based on the Quadruple Aim and Benefits Evaluation Frameworks: Scoping Review
J Med Internet Res 2020;22(12):e24568
URL: https://www.jmir.org/2020/12/e24568
doi:10.2196/24568
PMID:33289677

Original Paper

Electronic Cognitive Screen Technology for Screening Older Adults With Dementia and Mild Cognitive Impairment in a Community Setting: Development and Validation Study

Joyce Y C Chan[1*], MPH; Adrian Wong[1,2*], PhD; Brian Yiu[1*], BBA; Hazel Mok[1*], BSc; Patti Lam[3*], MSc; Pauline Kwan[1*], MSSc; Amany Chan[3*], MPhil; Vincent C T Mok[1,2,4*], MD; Kelvin K F Tsoi[5*], PhD; Timothy C Y Kwok[1,2,3*], MD

[1] Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[2] Therese Pei Fong Chow Research Centre for Prevention of Dementia, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[3] Jockey Club Centre for Osteoporosis Care and Control, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[4] Gerald Choa Neuroscience Centre, Lui Che Woo Institute of Innovative Medicine, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[5] Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[*] all authors contributed equally

Corresponding Author:
Timothy C Y Kwok, MD
Department of Medicine and Therapeutics
The Chinese University of Hong Kong
9/F., Lui Che Woo Clinical Sciences Building, Prince of Wales Hospital
30-32 Ngan Shing St, Shatin, N.T.
Hong Kong, 852
Hong Kong
Phone: 852 3505 3145
Email: tkwok@cuhk.edu.hk

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/1/e26724/

Abstract

Background: A digital cognitive test can be a useful and quick tool for the screening of cognitive impairment. Previous studies have shown that the diagnostic performance of digital cognitive tests is comparable with that of conventional paper-and-pencil tests. However, the use of commercially available digital cognitive tests is not common in Hong Kong, which may be due to the high cost of the tests and the language barrier. Thus, we developed a brief and user-friendly digital cognitive test called the Electronic Cognitive Screen (EC-Screen) for the detection of mild cognitive impairment (MCI) and dementia of older adults.

Objective: The aim of this study was to evaluate the performance of the EC-Screen for the detection of MCI and dementia in older adults.

Methods: The EC-Screen is a brief digital cognitive test that has been adapted from the Rapid Cognitive Screen test. The EC-Screen uses a cloud-based platform and runs on a tablet. Participants with MCI, dementia, and cognitively healthy controls were recruited from research clinics and the community. The outcomes were the performance of the EC-Screen in distinguishing participants with MCI and dementia from controls, and in distinguishing participants with dementia from those with MCI and controls. The cohort was randomly split into derivation and validation cohorts based on the participants' disease group. In the derivation cohort, the regression-derived score of the EC-Screen was calculated using binomial logistic regression. Two predictive models were produced. The first model was used to distinguish participants with MCI and dementia from controls, and the second model was used to distinguish participants with dementia from those with MCI and controls. Receiver operating characteristic curves were constructed and the areas under the curves (AUCs) were calculated. The performances of the two predictive models were tested using the validation cohorts. The relationship between the EC-Screen and paper-and-pencil Montreal Cognitive Assessment-Hong Kong version (HK-MoCA) was evaluated by the Pearson correlation coefficient.

Results: A total of 126 controls, 54 participants with MCI, and 63 participants with dementia were included in the study. In differentiating participants with MCI and dementia from controls, the AUC of the EC-Screen in the derivation and validation cohorts was 0.87 and 0.84, respectively. The optimal sensitivity and specificity in the derivation cohorts were 0.81 and 0.80, respectively. In differentiating participants with dementia from those with MCI and controls, the AUC of the derivation and validation cohorts was 0.90 and 0.88, respectively. The optimal sensitivity and specificity in the derivation cohort were 0.83 and 0.83, respectively. There was a significant correlation between the EC-Screen and HK-MoCA (r=–0.67, P<.001).

Conclusions: The EC-Screen is suggested to be a promising tool for the detection of MCI and dementia. This test can be self-administered or assisted by a nonprofessional staff or family member. Therefore, the EC-Screen can be a useful tool for case finding in primary health care and community settings.

(J Med Internet Res 2020;22(12):e17332) doi:10.2196/17332

KEYWORDS

EC-Screen; cognitive screening; dementia; mild cognitive impairment

Introduction

Dementia is a global challenge due to the aging population. The prevalence of dementia in older adults ranges from 5% to 7%, and the prevalence of mild cognitive impairment (MCI) ranges from 10% to 20% [1-3]. There are almost 10 million new cases of dementia diagnosed every year worldwide [4]. Studies have shown that early treatment and intervention can help to slow down cognitive decline in older adults [5-7]. The use of a cognitive screening test can facilitate early diagnosis, which in turn helps older adults with dementia and their families to work out a short-term coping and long-term care plan so that they can receive proper dementia-related care, advice, and support in a timely manner, and can live in the community. Improved community support can help to delay or reduce reliance on high-cost residential care services [8].

Paper-and-pencil cognitive screening tests such as the Montreal Cognitive Assessment (MoCA) are commonly used for the detection of cognitive impairment [9]. Although the utility of paper-and-pencil cognitive tests is generally good [10,11], most of these tests must be administered by professional staff, which increases the waiting time for patients and also risks introducing rater biases in test administration and scoring. Moreover, the calculation of the cutoff scores in paper-and-pencil cognitive tests cannot take behavioral data such as response time into account. Furthermore, “practice effects” occur with repeated applications, which could undermine the usefulness of the tests for measuring either the treatment response or the monitoring of disease progression [12,13]. In addition, older adults may not be motivated to seek out or undergo cognitive assessment with health care professionals, or they may have difficulties in accessing health care services. Indeed, a meta-analysis reported a high rate of undetected dementia, especially in China and India [14]. Therefore, it is important to find a way to help family members and health care professionals decide whether it is necessary to seek professional assessment by detecting early signs of cognitive impairment in the older people in their lives or in their care.

Recent studies have proposed the use of digital cognitive tests to overcome some of the above-mentioned barriers, as digital cognitive tests provide automatic, standardized administration procedures, including presentation of the stimulus, scoring, and performance classification [15-18]. Digital cognitive tests can be self-administered or used with minimal assistance by family members or nonprofessional staff, which can significantly increase access to cognitive screening in the general community. Furthermore, digital cognitive tests allow for accurate measurement of participants’ response time, which is known to be affected at an early stage in cognitive disorders [19]. A previous study showed that the diagnostic performance of digital cognitive tests is comparable with that of traditional paper-and-pencil tests [20]. Therefore, digital cognitive tests may play a helpful role as a preliminary screen in the workflow of cognitive assessment; those who show deficits may then undergo further assessment by professionals, thus facilitating better health care resource utilization. However, despite the availability of some commercial digital cognitive tests, their use is not common in Hong Kong, which may be due to the high cost of the tests and language barrier. Therefore, there is a need to develop a brief, user-friendly, and inexpensive digital cognitive test.

To fill this gap, we developed a brief digital cognitive test called the Electronic Cognitive Screen (EC-Screen) using a cloud-based platform that runs on a tablet. The EC-Screen is adapted from the Rapid Cognitive Screen (RCS), which is a short and well-validated paper-and-pencil cognitive test [21]. We aim to promote the use of the EC-Screen in primary health care and community settings in Hong Kong, such as in general practitioner clinics and community elder centers. The objective of this study was to evaluate the validity and performance of the EC-Screen for the detection of MCI and dementia in older adults.

Methods

Recruitment of Participants

This study was approved by the clinical research ethics committee of the Chinese University of Hong Kong (CUHK). Participants were recruited from research clinics of the Geriatrics and Neurology divisions of the CUHK, and from a community elderly center, namely the Jockey Club Center for Positive Aging in Hong Kong. The recruitment period was from March to November 2019. The inclusion criteria of the participants were aged ≥55 years, able to communicate in the Chinese language, and adequate perceptual-motor ability so as to be able to participate in cognitive testing. The exclusion criteria were people with uncontrolled psychiatric illnesses and participants who selected the illiterate version. This is because the answer time in the illiterate version is longer than that of the standard version, and therefore we excluded this version for the analysis to enable effective comparison.

Participants with MCI and dementia were consecutive patients from the Geriatrics and Neurology divisions of the CUHK. All of the participants with a diagnosis of MCI and dementia were assessed by a geriatrician. The diagnostic criterion of MCI was based on the Petersen criterion [22], and the diagnostic criterion of dementia was based on the International Classification of Diseases version 10 [23]. The healthy controls were recruited from the Jockey Club Center for Positive Aging and the Division of Neurology of the CUHK. The controls underwent neuropsychological assessment and the Hong Kong version of the MoCA (HK-MoCA) [9,24], and were assessed as cognitively healthy. A purposive sampling method was used [25]. All of the participants provided informed consent through the EC-Screen platform on a designated page that was designed to obtain consent, and the participants also signed a written consent form agreeing to participate in the research.

EC-Screen

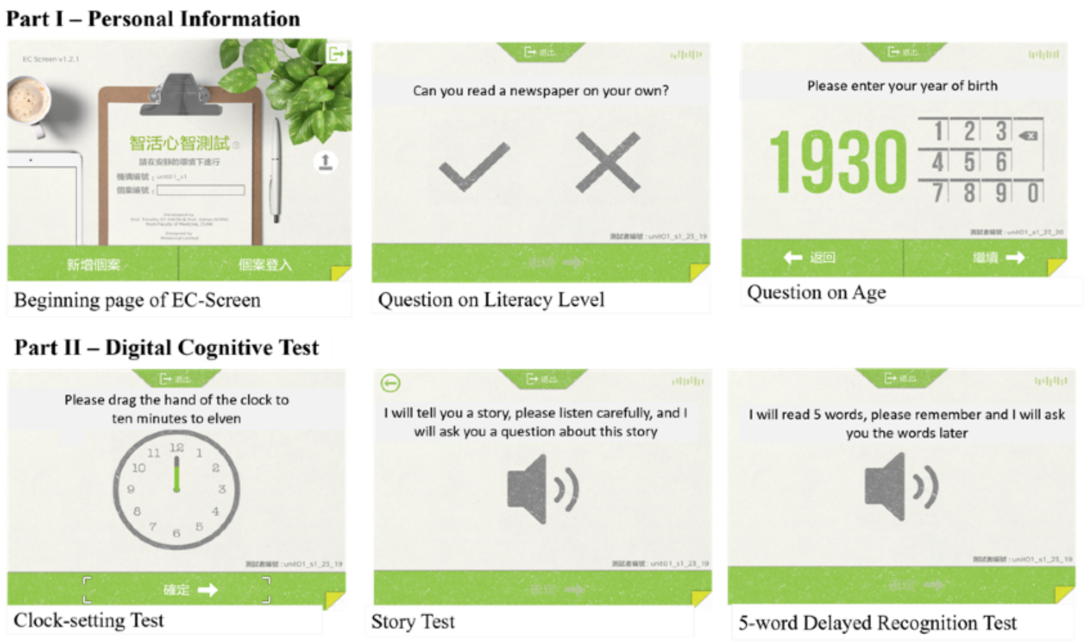
The EC-Screen was developed by the Department of Medicine and Therapeutics and the Division of Neurology of the CUHK. A local software company assisted with software development. The platform reads out all of the questions and then prompts the participant to select the answer on the touchscreen. Modifications of the test instructions were made for any illiterate participant by providing the option for the questions and possible responses to be read out by the software. For participants who are able to read, the platform only reads out the questions and the response choices are shown on the screen. The administration time of the EC-Screen is approximately 5 minutes.

The EC-Screen is composed of two parts. The first part collects the participant's personal information and the second part is the digital cognitive test. In the personal information part, participants are required to enter their year of birth, gender, education level, and area of residence. In the digital cognitive test part, the participants are required to answer three subtests, including a clock-setting test, story test, and 5-word delayed recognition test. The clock-setting test assesses executive functions and visuospatial abilities, the story test assesses mental flexibility, and the 5-word delayed recognition test assesses memory function. The digital cognitive test part of the EC-Screen was adapted from the RCS [21]. The Chinese Cantonese translation of the RCS was provided by Saint Louis University (St. Louis, MO, USA). The Chinese translation of the test items and the content validity were reviewed by a team of experienced experts, including a geriatrician (TK) and a clinical psychologist (AW). A comparison between RCS and EC-Screen is shown in Multimedia Appendix 1.

In brief, the flow of the digital cognitive test part of the EC-Screen is as follows: At the beginning of the test, the participant is required to learn 5 two-syllable words that are read out by the platform. The participant is then required to answer the clock-setting test in which a clock is presented on the screen and the participant is required to set the clock hands to a specified time. After the clock-setting test, the participant is required to answer the story test, which is a story-based fact conversion test. The platform reads a short story and the participant is required to remember the details of the story and appropriately identify the fact that a well-known landmark belongs to a certain region in Hong Kong. After the story test, the participant is required to answer the 5-word delayed recognition test. A total of 12 two-syllable words containing 5 target words and 7 distractors are presented on the screen, and the participant is asked to indicate which 5 words are the target word that he/she learned at the beginning of the test.

The flow of the EC-Screen is shown in Multimedia Appendix 2 and associated screenshots are shown in Figure 1. The total score of the delayed recognition test is 5, and the scoring of the clock-setting test and story test is simply dichotomized as correct or incorrect. The measurement of the time score in each subtest is focused on the interval between the end of the presentation of the instructions and the completion of the task.

Figure 1. Screenshots of EC-Screen.



Beginning page of EC-Screen

Can you read a newspaper on your own?

Question on Literacy Level

Please enter your year of birth

Question on Age

Please drag the hand of the clock to ten minutes to elven

Clock-setting Test

I will tell you a story, please listen carefully, and I will ask you a question about this story

Story Test

I will read 5 words, please remember and I will ask you the words later

5-word Delayed Recognition Test

Comparison Test

The HK-MoCA was used as a comparison test. The HK-MoCA is a well-validated multidomain paper-and-pencil cognitive test that assesses visuospatial and executive functions, naming, memory, attention, abstraction, and orientation [9,24]. The maximum total score of the HK-MoCA is 30 and the administration time is approximately 10 to 15 minutes. The EC-Screen and HK-MoCA were administrated by trained research assistants. Both tests were conducted on the same day in the research clinic or elderly center. The assessors were not blind to the participants' clinical diagnosis during the implementation of the tests because the participants were patients in the clinics. However, since the administration of the EC-Screen is given automatically by the system, the results obtained with the EC-Screen are not affected by this lack of blinding.

Outcomes

The outcomes were the performance of the EC-Screen in distinguishing the participants with MCI and dementia from controls and in distinguishing the participants with dementia from those with MCI and controls.

Sample Size Calculation

Based on the estimated prevalence rates of dementia and MCI at 8.9% and 8.5%, respectively [26], with a power of 0.8, a type I error of 0.05, and an expected sensitivity and specificity of 0.80 each, it was determined that a minimal overall sample size of 108 was required in the derivation cohort [27].

Statistical Analysis

The cohort was randomly split into derivation and validation cohorts (6:4) according to the participants' disease group (ie, control, MCI, and dementia). In the derivation cohort, multivariable binary logistic regression analysis was performed to test the association between the individual scores and the time spent on each subtest. A list of variables was identified and preliminarily tested using the general linear model. Raw scores and the time spent on each subtest were selected for further testing in the regression model. The scores and time spent on the subtests were standardized to a z-score for analysis. Receiver operating characteristic (ROC) curves were constructed to examine the ability of the predicted probability derived from the optimal logistic regression model with the scores and time spent on the subtests of the EC-Screen used as the explanatory variables. Two predictive models were produced. The first predictive model was used to distinguish participants with MCI and dementia from controls, and the second predictive model was used to distinguish participants with dementia from those with MCI and controls. The areas under the ROC curves (AUCs) were calculated with a 95% CI. A cut-off point was derived at an optimal balance of sensitivity and specificity. The performances of the two predictive models were then tested using the validation cohort. Concurrent validity was evaluated based on the Pearson correlation coefficient between the predicted probability score of the EC-Screen and the total score of the HK-MoCA. A *P* value ≤.05 was regarded as statistically significant. Statistical analyses were conducted using R with the readxl, pROC, and ggpubr packages.

Results

Characteristics of Participants

A total of 283 participants were recruited for this study. However, data of 18 participants, including 6 controls, 5 with MCI, and 7 with dementia, failed to upload to the database platform due to technical problems. In addition, 22 participants, including 1 control, 9 with MCI, and 12 with dementia, used the illiterate version, and were therefore excluded from the analysis. As a result, the screening results of 243 participants, including 126 controls, 54 participants with MCI, and 63 participants with dementia, were analyzed (Figure 2). The participants were randomly split into the derivation cohort and validation cohort. The characteristics of the participants are shown in Table 1. There were no significant differences in age, gender, educational level, and HK-MoCA score between the derivation and validation cohorts.

Figure 2. Flow diagram of participants.

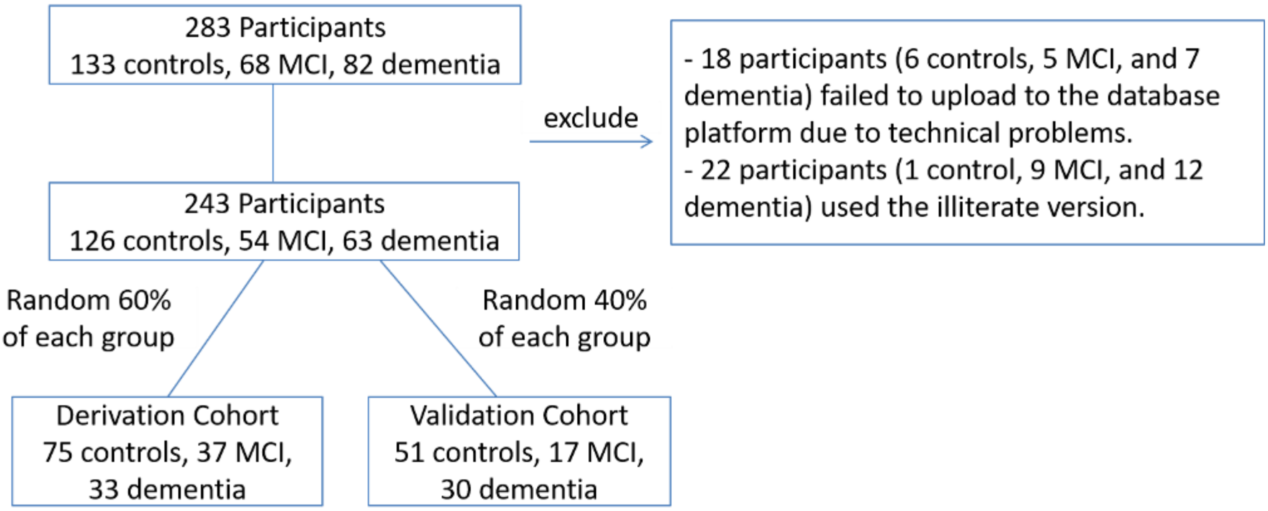


Table 1. Characteristics of the participants.

Characteristic	Derivation cohort			Validation cohort		
	Control (n=75)	MCI[a] (n=37)	Dementia (n=33)	Control (n=51)	MCI (n=17)	Dementia (n=30)
Mean age (years), mean (SD)	70.2 (8.1)	76.1 (6.3)	78.6 (7.1)	70.0 (9.3)	78.4 (7.6)	79.8 (7.5)
Females, n (%)	60 (80)	19 (51)	26 (79)	40 (78)	12 (71)	20 (67)
Education level, n (%)						
Primary level or below	29 (39)	18 (49)	18 (55)	13 (25)	12 (71)	23 (77)
Secondary level or above	46 (61)	19 (51)	15 (45)	38 (75)	5 (29)	7 (23)
HK-MoCA[b] score, mean (SD)	24.4 (3.3)	20.1 (3.7)	12.2 (6.4)	23.9 (3.9)	19.7 (4.4)	13.8 (5.1)
Z-score of EC-Screen[c] subtests, mean (SD)						
Clock-setting test	1.24 (1.5)	1.22 (1.5)	0.73 (1.3)	1.41 (1.5)	0.88 (1.4)	1.40 (1.5)
Delayed recognition test	3.52 (1.4)	2.97 (1.3)	1.45 (1.2)	3.67 (1.4)	2.59 (1.5)	1.43 (1.5)
Story test	1.08 (1.5)	0.81 (1.4)	0.55 (1.2)	1.18 (1.5)	1.59 (1.5)	0.90 (1.4)
Time spent on EC-Screen subtests (seconds), mean (SD)						
Clock-setting test	31.7 (13.9)	39.6 (16.8)	68.4 (51.3)	31.6 (12.4)	49.4 (27.6)	67.5 (42.0)
Delayed recognition test	28.0 (12.4)	36.3 (15.7)	43.8 (24.2)	30.3 (14.9)	48.2 (26.2)	55.6 (34.0)
Story test	75.3 (11.9)	75.8 (18.3)	89.1 (23.4)	73.7 (10.2)	84.7 (13.7)	87.6 (24.3)

[a] MCI: mild cognitive impairment.

[b] HK-MoCA: Hong Kong version of Montreal Cognitive Assessment.

[c] EC-Screen: Electronic Cognitive Screen.

Validation Results

Predictive Model for Distinguishing Participants With MCI and Dementia From Controls

A predicted probability score for having MCI and dementia derived from the EC-Screen was obtained by taking the raw score of the 5-word delayed recognition test and the clock-setting test, as well as the time spent in the 5-word delayed recognition test and clock-setting test into the following regression formula (see Multimedia Appendix 3 for the regression coefficients):

$$\text{logit}(p) = -1.015 - 0.08\,(\text{clock score}) - 0.68\,(\text{delayed recognition score}) + 0.03\,(\text{clock time}) + 0.05\,(\text{delayed recognition time})$$

In differentiating participants with MCI and dementia from controls, the AUCs of the EC-Screen in the derivation and validation cohort were high (Table 2). The optimal sensitivity and specificity in derivation cohort were 0.81 and 0.80, respectively, with the cut-off point of the predicted probability score identified as ≥0.43. The sensitivity and specificity in the validation cohort were equivalent but slightly lower. The ROC curves in the derivation cohort and the validation cohort did not show overfitting of the model (Figure 3).

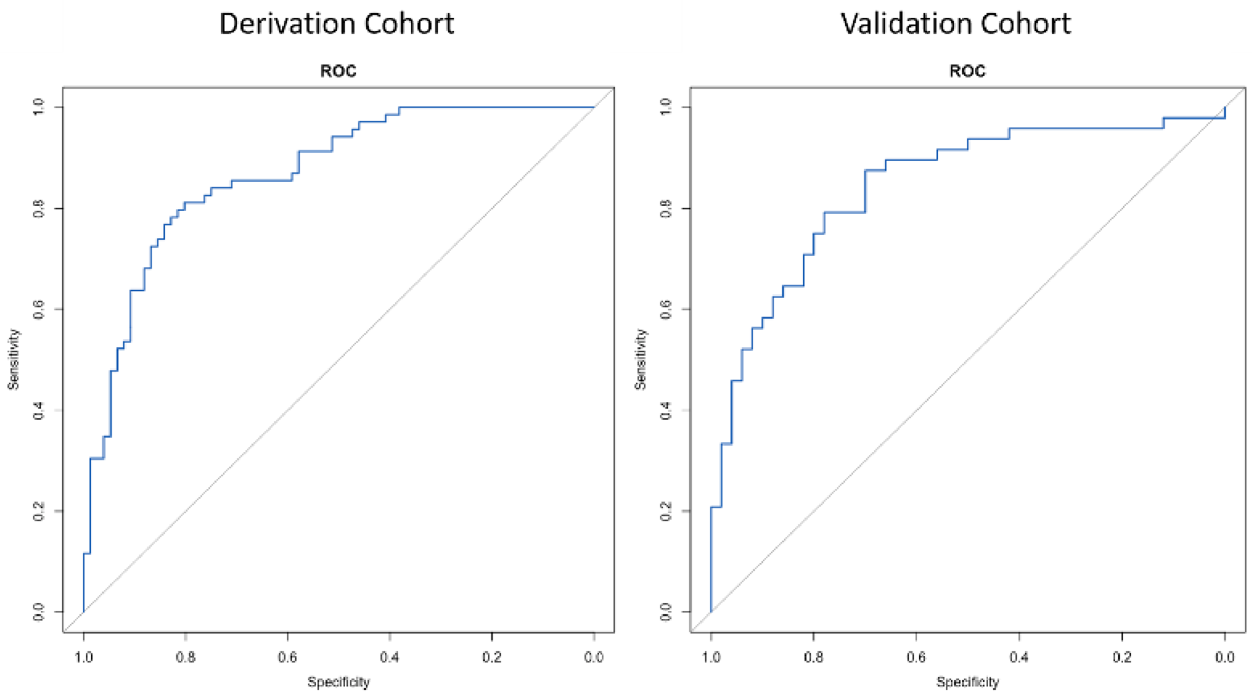
Table 2. Results of diagnostic performance of the EC-Screen–derived regression.

Regression model	AUC[a] (95% CI)	Cut-off value	Sensitivity	Specificity
Detect MCI[b] + dementia from controls				
Derivation cohort	0.87 (0.81-0.93)	0.43	0.81	0.80
Validation cohort	0.84 (0.76-0.92)	0.43	0.79	0.78
Detect dementia from MCI + controls				
Derivation cohort	0.90 (0.84-0.95)	0.22	0.83	0.83
Validation cohort	0.88 (0.81-0.96)	0.22	0.82	0.76

[a] AUC: area under the curve.

[b] MCI: mild cognitive impairment.

Figure 3. Receiver operating characteristic (ROC) curves of EC-Screen for discriminating among participants with mild cognitive impairment and dementia.



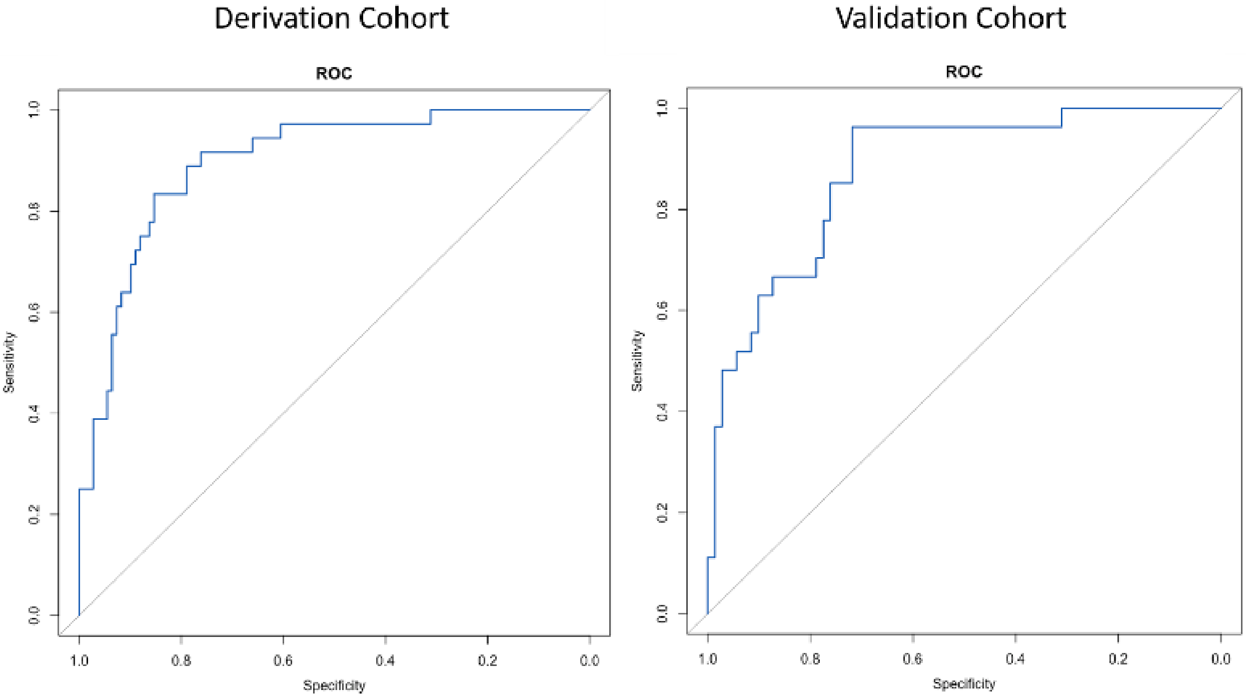
Predictive Model for Distinguishing Dementia Participants from MCI Participants and Controls

A predicted probability score for having dementia derived from the EC-Screen was obtained by taking the raw score of the 5-word delayed recognition test, the clock-setting test, and the story test, as well as the time spent in the 5-word delayed recognition test and clock-setting test into the following regression formula (see Multimedia Appendix 3 for regression coefficients):

$$\text{logit}(p) = -1.05 - 0.26\,(\textit{clock score}) - 0.15\,(\textit{story score}) - 0.87\,(\textit{delayed recognition score}) + 0.02\,(\textit{clock time}) + 0.03\,(\textit{delayed recognition time})$$

In differentiating dementia participants from MCI participants and controls, the AUCs of the EC-Screen of the derivation and validation cohorts were high (Table 2). The optimal sensitivity and specificity in the derivation cohort were 0.83 and 0.83, respectively, with the cut-off point of the predicted probability score identified as ≥0.22. The sensitivity and specificity of the validation cohort were equivalent, but slightly lower (Table 2). The ROC curves in the derivation cohort and the validation cohort did not show overfitting of the model (Figure 4).

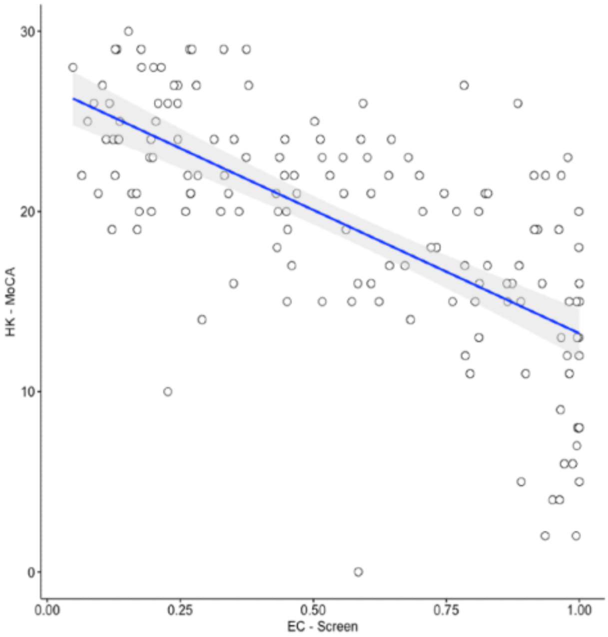
Figure 4. Receiver operating characteristic (ROC) curves of EC-Screen for discriminating among participants with dementia from those with mild cognitive impairment and controls in the derivation and validation cohorts.



Concurrent Validity

There was a significant correlation between the predicted probability score of the EC-Screen–derived regression and the total score of the HK-MoCA (*r*=–0.67, *P*<.001) (Figure 5).

Figure 5. Scatterplot showing the relationship between the predicted probability of EC-Screen–derived regression and the Hong Kong version of the Montreal Cognitive Assessment (HK-MoCA).



Administration Process

The average administration time of the EC-Screen was 4.5 minutes. In the qualitative report from users, some of the participants could complete the test by themselves and some of them required assistance from a helper to use the tablet. Therefore, the EC-Screen can be self-administered or assisted by a nonprofessional staff or family member.

Discussion

Principal Findings

This study shows that the EC-Screen has good criteria and concurrent validity to identify older adults with MCI and dementia. It is brief and only requires 5 minutes to administer. The automatic administration and scoring algorithm can ease the workload of professional and health care staff. The EC-Screen is a promising tool to use in community centers and primary health care clinics, and thus older adults at risk can receive cognitive screening promptly in the community.

Digital methods allow for capturing response times accurately, which are not easy to capture with paper-and-pencil tests. Participants with MCI and dementia require a longer time in the clock-setting test, which might reflect problems in processing speed and executive functions. The traditional method of the clock-drawing test requires more precise fine motor control, especially when drawing on a tablet. In contrast, the clock-setting test requires less fine motor control and is thus easier for older adults, which minimizes the confounding effects of physical constraints and poor motor dexterity. The diagnostic performance of the clock-setting test was slightly better than that determined for the clock-drawing test in a pilot study [28]. Therefore, the digital clock-setting test was used in the EC-Screen. We used a digital delayed recognition test instead of a traditional method of the delayed free recall test because current voice recognition technology is not yet able to automatically understand a stranger's voice correctly. Therefore, the digital version of the delayed recognition test was considered to be more practical for use in community settings than the digital version of the delayed free recall test. Delayed recognition tests are commonly used for cognitive screening. Previous studies have shown that the diagnostic performance of the digital version of delayed recognition tests effectively detected both MCI and dementia [20,29]. Both the raw score and time spent in the 5-word delayed recognition test were statistically significant in the regression model, indicating that problems in memory retrieval and processing efficiency are important markers of cognitive disorders. Impairment of delayed memory recognition could reflect problems in encoding, consolidation, or storage, which together comprise a cardinal feature of Alzheimer disease. Therefore, performance on delayed recognition may serve as a more sensitive marker than delayed free recall for identifying patients with early cognitive decline at risk for progression to Alzheimer disease [30]. The question of the story test is not the exact content of the story and requires the participant to convert a fact that a landmark belongs to a given district. Therefore, the story test can assess the mental ability to switch between two concepts.

Some older adults may be aware of their memory decline, and some family members may worry about their parents or grandparents with potential signs of cognitive decline; thus, the EC-Screen can help them to decide whether it is necessary to seek medical and professional advice. Older adults can obtain the cognitive screening assessment from community centers or elderly centers. Therefore, the EC-Screen can promote timely assessment for older adults at risk in the community.

A digital cognitive test can capture behavioral data accurately. Some studies found that drawing time or drawing process can be a predictive factor for cognitive impairment [31,32]. In the last decade, some multidomain digital cognitive tests have been investigated, including Brain on Track [13] and Computerized Cognitive Screen [33]. These digital cognitive tests assess memory, attention, processing speed, and executive functions, and the administration time is around 20 to 25 minutes. The AUC of Brain on Track for the detection of MCI was 0.86 and the AUC of the Computerized Cognitive Screen for the detection of MCI and dementia was 0.78. There are some commercially available computerized cognitive test batteries such as the Computer assessment for Mild Cognitive Impairment (CAMCI) [34] and CNS Vital Signs [35]. However, the administration time of these tests is longer (>20 minutes) and the costs are rather high. The EC-Screen showed comparable diagnostic performance with a simpler design. Further evaluation of the EC-Screen in a larger cohort of older people recruited from various sources in the community is ongoing.

Previous studies have reported that the performance on cognitive tests such as the MoCA is affected by education [36-38]. Such education effects are more obvious in Asian countries, as elderly people in these countries are generally less educated [36,38]. The EC-Screen has an adapted version with illiterate options for administration of the test. Twenty-two participants chose the illiterate version in this study. The design of the illiterate version is tailor-made for participants with a very low education level. However, the administration time of this version is longer than that of the standard version, and therefore we excluded the participants who used the illiterate version from this analysis. We are currently planning to separately analyze the participants who took the illiterate version when a larger sample is obtained.

Limitations

There were some limitations to this study. First, the sample size was small, and therefore EC-Screen results showing that it can discriminate among dementia, MCI, and controls need to be confirmed in a larger study. Second, the test/retest reliability was not investigated in this study and needs to be examined in the future.

Conclusions

The EC-Screen is suggested to be a promising tool for the detection of MCI and dementia. The EC-Screen is brief and can be self-administered or assisted by a nonprofessional staff or family member. Therefore, it can be a useful tool for case finding in primary health care and community settings.

Acknowledgments

We thank Professor JE Morley, Saint Louis University School of Medicine, USA, for having agreed to let us adapt his Rapid Cognitive Screen test into EC-Screen for older Chinese people. We are also grateful for the funding support from the Hong Kong Jockey Club Charities trust, and for the support of the Jockey Club Centre for Positive Ageing for its contribution in design and data collection. We appreciate the support from Mindvivid Limited in software development of EC-Screen. We also thank Ms. Anthea Ng, research assistant of the Division of Neurology at The Chinese University of Hong Kong, for her help in data collection and entry.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of Rapid Cognitive Screen and EC-Screen.
[DOCX File , 18 KB - jmir_v22i12e17332_app1.docx]

Multimedia Appendix 2
Description of the flow of EC-Screen.
[DOCX File , 23 KB - jmir_v22i12e17332_app2.docx]

Multimedia Appendix 3
Regression coefficients of the regression models.
[DOCX File , 24 KB - jmir_v22i12e17332_app3.docx]

References

1. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 2013 Jan;9(1):63-75. [doi: 10.1016/j.jalz.2012.11.007] [Medline: 23305823]
2. Petersen RC. Mild Cognitive Impairment. N Engl J Med 2011 Jun 09;364(23):2227-2234. [doi: 10.1056/nejmcp0910237]
3. Langa KM, Levine DA. The diagnosis and management of mild cognitive impairment: a clinical review. JAMA 2014 Dec 17;312(23):2551-2561 [FREE Full text] [doi: 10.1001/jama.2014.13806] [Medline: 25514304]
4. Dementia. World Health Organization. 2019 Sep 19. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2020-07-02]
5. Hill NTM, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized Cognitive Training in Older Adults With Mild Cognitive Impairment or Dementia: A Systematic Review and Meta-Analysis. Am J Psychiatry 2017 Apr 01;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
6. Karssemeijer EGA, Aaronson JA, Bossers WJ, Smits T, Olde Rikkert MGM, Kessels RPC. Positive effects of combined cognitive and physical exercise training on cognitive function in older adults with mild cognitive impairment or dementia: A meta-analysis. Ageing Res Rev 2017 Nov;40:75-83. [doi: 10.1016/j.arr.2017.09.003] [Medline: 28912076]
7. Guitar NA, Connelly DM, Nagamatsu LS, Orange JB, Muir-Hunter SW. The effects of physical exercise on executive function in community-dwelling older adults living with Alzheimer's-type dementia: A systematic review. Ageing Res Rev 2018 Nov;47:159-167. [doi: 10.1016/j.arr.2018.07.009] [Medline: 30102996]
8. Prince M, Bryce R, Ferri C. The benefits of early diagnosis and intervention. Alzheimer's Disease International World Alzheimer Report. 2011. URL: https://www.alz.co.uk/research/world-report-2011 [accessed 2020-04-18]
9. Nasreddine Z, Phillips N, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
10. Tsoi KKF, Chan JYC, Hirai HW, Wong A, Mok VCT, Lam LCW, et al. Recall Tests Are Effective to Detect Mild Cognitive Impairment: A Systematic Review and Meta-analysis of 108 Diagnostic Studies. J Am Med Dir Assoc 2017 Sep 01;18(9):807. [doi: 10.1016/j.jamda.2017.05.016] [Medline: 28754516]
11. Tsoi KKF, Chan JYC, Hirai HW, Wong SYS, Kwok TCY. Cognitive Tests to Detect Dementia: A Systematic Review and Meta-analysis. JAMA Intern Med 2015 Sep;175(9):1450-1458. [doi: 10.1001/jamainternmed.2015.2152] [Medline: 26052687]
12. Sternin A, Burns A, Owen AM. Thirty-Five Years of Computerized Cognitive Assessment of Aging-Where Are We Now? Diagnostics (Basel) 2019 Sep 06;9(3):114 [FREE Full text] [doi: 10.3390/diagnostics9030114] [Medline: 31489940]
13. Ruano L, Severo M, Sousa A, Ruano C, Branco M, Barreto R, et al. Tracking Cognitive Performance in the General Population and in Patients with Mild Cognitive Impairment with a Self-Applied Computerized Test (Brain on Track). J Alzheimers Dis 2019;71(2):541-548. [doi: 10.3233/JAD-190631] [Medline: 31424407]

14. Lang L, Clifford A, Wei L, Zhang D, Leung D, Augustine G, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open 2017 Feb 03;7(2):e011146 [FREE Full text] [doi: 10.1136/bmjopen-2016-011146] [Medline: 28159845]
15. Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: a systematic review. Alzheimers Dement 2008 Nov;4(6):428-437 [FREE Full text] [doi: 10.1016/j.jalz.2008.07.003] [Medline: 19012868]
16. Tierney MC, Lermer MA. Computerized cognitive assessment in primary care to identify patients with suspected cognitive impairment. J Alzheimers Dis 2010;20(3):823-832. [doi: 10.3233/JAD-2010-091672] [Medline: 20413868]
17. Zygouris S, Tsolaki M. Computerized cognitive testing for older adults: a review. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):13-28 [FREE Full text] [doi: 10.1177/1533317514522852] [Medline: 24526761]
18. de Oliveira MO, Brucki SMD. Computerized Neurocognitive Test (CNT) in mild cognitive impairment and Alzheimer's disease. Dement Neuropsychol 2014;8(2):112-116 [FREE Full text] [doi: 10.1590/S1980-57642014DN82000005] [Medline: 29213891]
19. Müller S, Preische O, Heymann P, Elbing U, Laske C. Increased Diagnostic Accuracy of Digital vs. Conventional Clock Drawing Test for Discrimination of Patients in the Early Course of Alzheimer's Disease from Cognitively Healthy Individuals. Front Aging Neurosci 2017;9:101. [doi: 10.3389/fnagi.2017.00101] [Medline: 28443019]
20. Chan JYC, Kwong JSW, Wong A, Kwok TCY, Tsoi KKF. Comparison of Computerized and Paper-and-Pencil Memory Tests in Detection of Mild Cognitive Impairment and Dementia: A Systematic Review and Meta-analysis of Diagnostic Studies. J Am Med Dir Assoc 2018 Sep;19(9):748-756. [doi: 10.1016/j.jamda.2018.05.010] [Medline: 29921507]
21. Malmstrom TK, Voss VB, Cruz-Oliver DM, Cummings-Vaughn LA, Tumosa N, Grossberg GT, et al. The Rapid Cognitive Screen (RCS): A Point-of-Care Screening for Dementia and Mild Cognitive Impairment. J Nutr Health Aging 2015 Aug;19(7):741-744. [doi: 10.1007/s12603-015-0564-2] [Medline: 26193857]
22. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999 Mar 01;56(3):303-308. [doi: 10.1001/archneur.56.3.303] [Medline: 10190820]
23. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. Geneva: WHO; 1993. URL: https://www.who.int/classifications/icd/en/GRNBOOK.pdf [accessed 2020-04-18]
24. Wong A, Xiong YY, Kwan PWL, Chan AYY, Lam WWM, Wang K, et al. The validity, reliability and clinical utility of the Hong Kong Montreal Cognitive Assessment (HK-MoCA) in patients with cerebral small vessel disease. Dement Geriatr Cogn Disord 2009;28(1):81-87. [doi: 10.1159/000232589] [Medline: 19672065]
25. Timmins F. Nursing Research Generating and Assessing Evidence for Nursing Practice. Nurse Educ Pract 2013 Nov;13(6):e29. [doi: 10.1016/j.nepr.2013.04.001]
26. Lam LCW, Tam CWC, Lui VWC, Chan WC, Chan SSM, Wong S, et al. Prevalence of very mild and mild dementia in community-dwelling older Chinese people in Hong Kong. Int Psychogeriatr 2008 Feb;20(1):135-148. [doi: 10.1017/S1041610207006199] [Medline: 17892609]
27. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. J Clin Epidemiol 2005 Aug;58(8):859-862. [doi: 10.1016/j.jclinepi.2004.12.009] [Medline: 16018921]
28. Chan J, Wong A, Tsoi K, Kwok T. Development of Electronic Cognitive Screen for Screening of People with Cognitive Impairment. A pilot validation study. 2019 May 29 Presented at: International Symposium on Digital Health; May 29, 2019; Hong Kong SAR.
29. Kingsbury R, Pachana NA, Humphreys M, Tehan G, Byrne GJ. Utility of a Computerised Cognitive Screen in MCI and Depression in an Older Population. Austral J Rehab Couns 2015 Oct 15;16(1):14-26. [doi: 10.1375/jrc.16.1.14]
30. De Simone MS, Perri R, Fadda L, Caltagirone C, Carlesimo GA. Predicting progression to Alzheimer's disease in subjects with amnestic mild cognitive impairment using performance on recall and recognition tests. J Neurol 2019 Jan;266(1):102-111. [doi: 10.1007/s00415-018-9108-0] [Medline: 30386876]
31. Tsoi KK, Leung NW, Lam MW. P2-184: An Automated Electronic Rating System for the Overlapped Pentagons in Dementia Screening Tests. Alzheimers Dement 2016 Jul 01;12(7S_Part_14):P687-P688. [doi: 10.1016/j.jalz.2016.06.1351]
32. Robens S, Heymann P, Gienger R, Hett A, Müller S, Laske C, et al. The Digital Tree Drawing Test for Screening of Early Dementia: An Explorative Study Comparing Healthy Controls, Patients with Mild Cognitive Impairment, and Patients with Early Dementia of the Alzheimer Type. J Alzheimers Dis 2019;68(4):1561-1574. [doi: 10.3233/JAD-181029] [Medline: 30909229]
33. Wong A, Fong C, Mok VC, Leung K, Tong RK. Computerized Cognitive Screen (CoCoSc): A Self-Administered Computerized Test for Screening for Cognitive Impairment in Community Social Centers. J Alzheimers Dis 2017;59(4):1299-1306. [doi: 10.3233/JAD-170196] [Medline: 28731437]
34. Saxton J, Morrow L, Eschman A, Archer G, Luther J, Zuccolotto A. Computer assessment of mild cognitive impairment. Postgrad Med 2009 Mar 13;121(2):177-185 [FREE Full text] [doi: 10.3810/pgm.2009.03.1990] [Medline: 19332976]
35. Gualtieri CT, Johnson LG. Neurocognitive testing supports a broader concept of mild cognitive impairment. Am J Alzheimers Dis Other Demen 2005;20(6):359-366 [FREE Full text] [doi: 10.1177/153331750502000607] [Medline: 16396441]
36. Lee J, Lee DW, Cho S, Na DL, Jeon HJ, Kim S, et al. Brief screening for mild cognitive impairment in elderly outpatient clinic: validation of the Korean version of the Montreal Cognitive Assessment. J Geriatr Psychiatry Neurol 2008 Jun;21(2):104-110. [doi: 10.1177/0891988708316855] [Medline: 18474719]

37. Pinto TCC, Machado L, Bulgacov TM, Rodrigues-Júnior AL, Costa MLG, Ximenes RCC, et al. Influence of Age and Education on the Performance of Elderly in the Brazilian Version of the Montreal Cognitive Assessment Battery. Dement Geriatr Cogn Disord 2018;45(5-6):290-299. [doi: 10.1159/000489774] [Medline: 29996142]
38. Wong A, Law LSN, Liu W, Wang Z, Lo ESK, Lau A, et al. Montreal Cognitive Assessment: One Cutoff Never Fits All. Stroke 2015 Dec;46(12):3547-3550. [doi: 10.1161/STROKEAHA.115.011226] [Medline: 26470772]

Abbreviations

AUC: area under the curve
CUHK: Chinese University of Hong Kong
EC-Screen: Electronic Cognitive Screen
HK-MoCA: Hong Kong version of the Montreal Cognitive Assessment
MCI: mild cognitive impairment
MoCA: Montreal Cognitive Assessment
RCS: Rapid Cognitive Screen
ROC: receiver operating characteristic

Edited by G Eysenbach; submitted 07.12.19; peer-reviewed by E Siegmund-Schultze, MDG Pimentel, K Baudin, D Scharre; comments to author 30.03.20; revised version received 21.07.20; accepted 31.10.20; published 18.12.20.

Please cite as:
Chan JYC, Wong A, Yiu B, Mok H, Lam P, Kwan P, Chan A, Mok VCT, Tsoi KKF, Kwok TCY
Electronic Cognitive Screen Technology for Screening Older Adults With Dementia and Mild Cognitive Impairment in a Community Setting: Development and Validation Study
J Med Internet Res 2020;22(12):e17332
URL: http://www.jmir.org/2020/12/e17332/
doi:10.2196/17332
PMID:33337341

Original Paper

Simulation Game Versus Multiple Choice Questionnaire to Assess the Clinical Competence of Medical Students: Prospective Sequential Trial

Tristan Fonteneau[1], MD; Elodie Billion[1], MD; Cindy Abdoul[1], MD; Sebastien Le[1], MD; Alice Hadchouel[1,2], MD, PhD; David Drummond[1,2], MD, PhD

[1]Department of Paediatric Pulmonology and Allergology, University Hospital Necker-Enfants Malades, Assistance Publique - Hôpitaux de Paris, Paris, France

[2]iLumens Simulation Department, University of Paris, Paris, France

Corresponding Author:
David Drummond, MD, PhD
Department of Paediatric Pulmonology and Allergology
University Hospital Necker-Enfants Malades
Assistance Publique - Hôpitaux de Paris
149 rue de Sèvres
Paris,
France
Phone: 33 1 44 49 48 48
Email: david.drummond@aphp.fr

Abstract

Background: The use of simulation games (SG) to assess the clinical competence of medical students has been poorly studied.

Objective: The objective of this study was to assess whether an SG better reflects the clinical competence of medical students than a multiple choice questionnaire (MCQ).

Methods: Fifth-year medical students in Paris (France) were included and individually evaluated on a case of pediatric asthma exacerbation using three successive modalities: high-fidelity simulation (HFS), considered the gold standard for the evaluation of clinical competence, the SG *Effic'Asthme*, and an MCQ designed for the study. The primary endpoint was the median kappa coefficient evaluating the correlation of the actions performed by the students between the SG and HFS modalities and the MCQ and HFS modalities. Student satisfaction was also evaluated.

Results: Forty-two students were included. The actions performed by the students were more reproducible between the SG and HFS modalities than between the MCQ and HFS modalities (P=.04). Students reported significantly higher satisfaction with the SG (P<.01) than with the MCQ modality.

Conclusions: The SG *Effic'Asthme* better reflected the actions performed by medical students during an HFS session than an MCQ on the same asthma exacerbation case. Because SGs allow the assessment of more dimensions of clinical competence than MCQs, they are particularly appropriate for the assessment of medical students on situations involving symptom recognition, prioritization of decisions, and technical skills.

Trial Registration: ClinicalTrials.gov NCT03884114; https://clinicaltrials.gov/ct2/show/NCT03884114

(J Med Internet Res 2020;22(12):e23254) doi:10.2196/23254

KEYWORDS

serious game; simulation game; assessment; professional competence; asthma; pediatrics

Introduction

An essential mission of medical schools is to regularly assess the clinical competence of their medical students. These assessments are made difficult by the multidimensional aspects of medical competence, and different methods have been developed [1]. High-fidelity simulation (HFS) and objective structured clinical examination (OSCE) are currently considered the best modalities to assess the clinical competence of medical students because they represent reliable, valid, and acceptable assessment methods without any risk for patients [2-4]. In addition, they allow several dimensions of medical competence to be assessed at the same time, such as knowledge, clinical reasoning, technical skills, and teamwork. However, HFS and OSCE are difficult to implement on a large scale because they require a lot of human resources and faculty time is a scarce resource [5]. In addition, summative assessments with these techniques are suboptimal because of the subjectivity of the evaluators even when a standardized checklist is used and because of the difficulty of proposing the same situations to all medical students, who cannot be evaluated at the same time. Alternatively, multiple choice questionnaires (MCQ) represent a reliable, acceptable, and inexpensive method of assessment that can be widely deployed and offer automated correction. However, MCQs have several drawbacks: they only test knowledge and clinical reasoning, they create situations in which a student can answer a question by recognizing the correct option but would not have been able to answer it in the absence of options (cueing effect), and they limit choices to 4 or 5 proposals when there are many more in real life [4]. There is, therefore, a need for an interim evaluation method, bringing together most of the strengths of the HFS/OSCE and MCQ modalities.

Simulation games (SGs) may represent this solution. SGs combine the features of serious games and simulation. Serious games are defined as games specifically designed for a serious purpose, such as providing health professions education [6]. As serious games, SGs incorporate rules and predefined educational objectives to win the game. This is different from virtual simulators, which can be used without predefined objectives. For example, in Microsoft Flight Simulator, the simulator reproduces the conditions of a real environment but no objective is provided to the player, who can choose the airport they want to fly to [7]. SGs can be defined as a type of serious games designed to closely simulate real-world activities [8]. They belong to the broader group of serious games in that they include preestablished objectives (eg, the patient's recovery). They share with virtual simulators their realistic, artificial environment in which learners can apprehend the consequences of their decisions.

As serious games, SGs promote attention, active learning, feedback, and consolidation, which have been identified as the four main pillars of learning by cognitive scientists [9]. Moreover, the objective of winning incorporated in SGs enhances learners' motivation and engagement [7,10]. As simulations, SGs allow users to acquire complex behavioral and technical skills that cannot be entirely acquired through knowledge-based training methods alone and also have the advantage of being risk free for patients and learners [11,12].

SGs have become increasingly popular in the training of health professionals in recent years, and their educational effectiveness has been confirmed by several studies [8,13,14]. However, they have been poorly evaluated as an assessment tool for medical students [15]. Similar to MCQs, SGs offer a standardized assessment that can be given simultaneously on a large scale and inexpensive, automated, and objective correction. They go beyond MCQs by allowing a larger degree of freedom in the options that can be chosen by students and assessment of certain characteristics of the physical examination (inspection, auscultation) and technical skills such as the use of a pressurized metered-dose inhaler with a spacer. However, it is unclear whether an SG is a better method of assessment than a conventional MCQ for evaluating the clinical competence of medical students.

The objective of this study was to compare the performance of an SG and an MCQ to assess the clinical competence of medical students in a scenario of pediatric asthma exacerbation. The gold standard chosen was HFS, and the actions performed on the SG and the responses provided on the MCQ were compared with the actions performed by students during the HFS session to analyze the degree of concordance between the SG and HFS modalities on the one hand and the MCQ and HFS modalities on the other hand.

Methods

Ethics

A prospective, simulation-based trial was conducted in the department of simulation in health care iLumens (a multidisciplinary university medical laboratory focused on digital health education) in Paris Descartes University (France). The study was approved by the ethics committee of our institution (CENEM 2019-15-DD) and registered in ClinicalTrials.gov [NCT03884114].

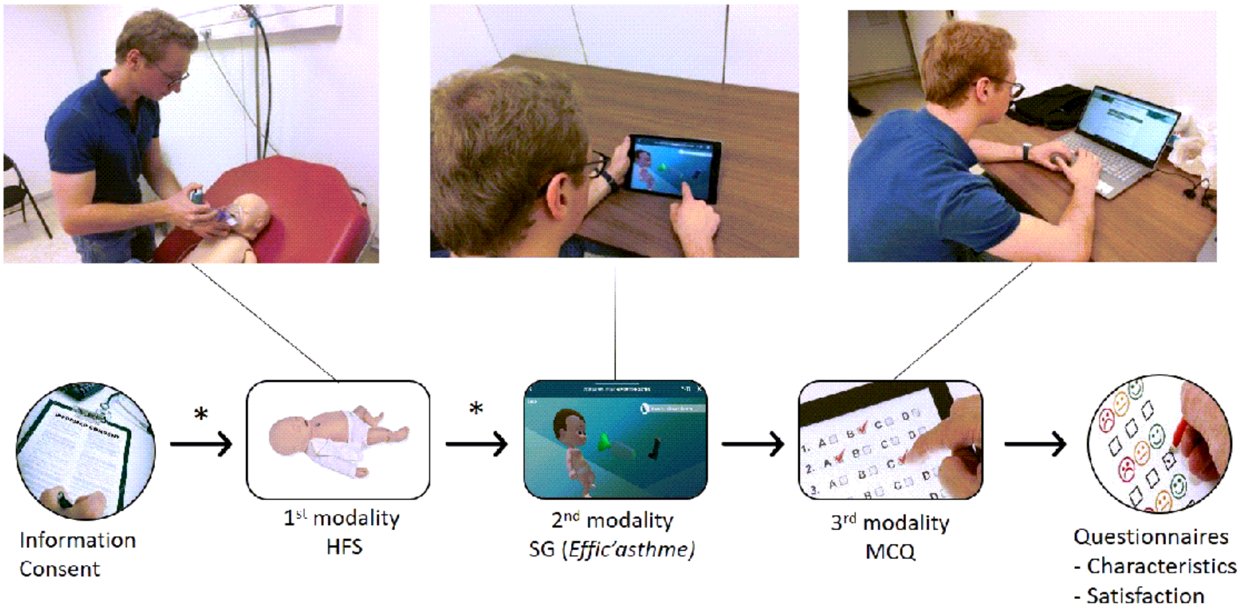
Participants

Participants were fifth-year medical students from the French faculty of medicine Paris Descartes who participated voluntarily. They were part of two classes and went to the department of simulation in March and June 2019 after their 3-month pediatrics courses which included a 2-hour course on pediatric asthma exacerbation management and within 15 days after their pediatrics exams. Recruitment was done by emails and during their pediatric exams. Written consent was obtained, and demographic data were collected from participants.

Study Design

The study design is presented in Figure 1. Participants were successively evaluated on the same scenario of a moderate asthma exacerbation occurring in a private medical practice using three different modalities (HFS, SG, and MCQ) for a maximum duration of 12 minutes for each modality. Participants started by watching a 5-minutes tutorial video on the features of the high-fidelity manikin SimBaby (Laerdal Medical AS) used and its environment. They were then individually evaluated on the management of a moderate asthma exacerbation scenario on HFS. No debriefing was provided, and participants were invited to watch a 3-minute video on the features of the SG (*Effic'Asthme*, iLumens, and Dowino). The second evaluation was conducted on the same scenario using the SG. Again, no debriefing was provided, and the third evaluation was conducted on the same scenario using an MCQ. Finally, participants completed a questionnaire on their characteristics, followed by a second questionnaire about their satisfaction with each evaluation modality and ending with an oral debriefing.

Figure 1. Study design. HFS: high fidelity simulation; SG: simulation game; MCQ: multiple choice questionnaire; *: tutorial video.

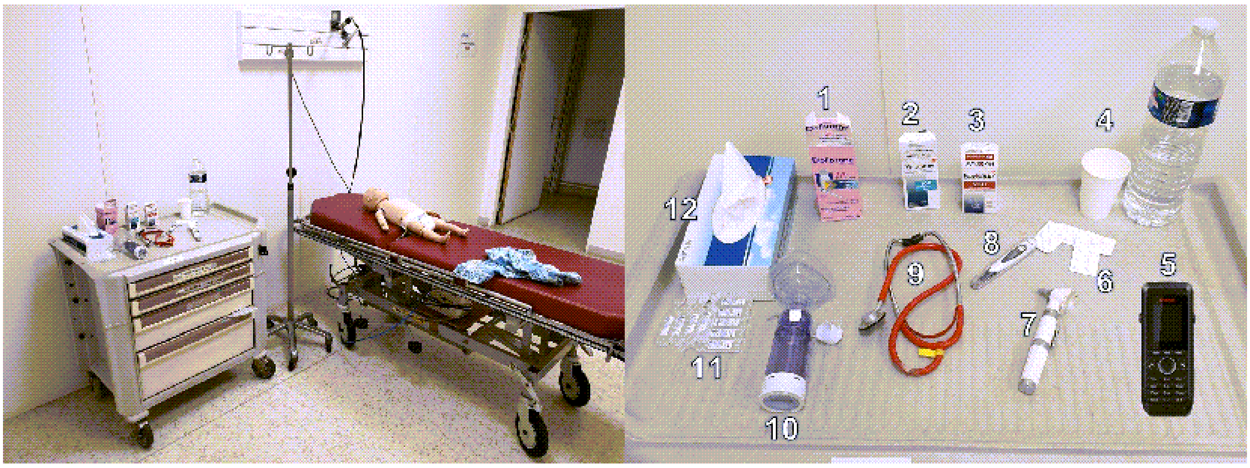


Scenario and Evaluation Modalities

A single scenario was used for all three modalities. It corresponded to a moderate exacerbation of asthma in a child aged 12-months who did not respond to short-acting beta-agonists (SABAs). A full description of the scenario is presented in Multimedia Appendix 1 (Table S1). Participants were expected to administer the emergency treatment (salbutamol, SABAs) to the child with the correct inhalation technique and dose, repeat this administration after no improvement was noted, repeat this administration again adding oral corticosteroids, and finally refer the child to the nearest pediatric emergency department.

The first evaluation modality studied was HFS. As it was impossible to assess the students on their ability to manage an asthma exacerbation with a real patient, we considered HFS the gold standard assessment method that would best reflect students' clinical competence. The Simbaby manikin used can reproduce many signs of an asthma exacerbation. In the simulation room, participants could use the same items as in the SG (Figure 2): an emergency treatment (salbutamol, SABAs), a controller treatment (fluticasone), an asthma spacer with a facial mask, oral steroids in tablets (prednisone), a glass of water, paracetamol, saline nose drops for nasal airway clearance, and a phone. All actions performed by the participants were video recorded.

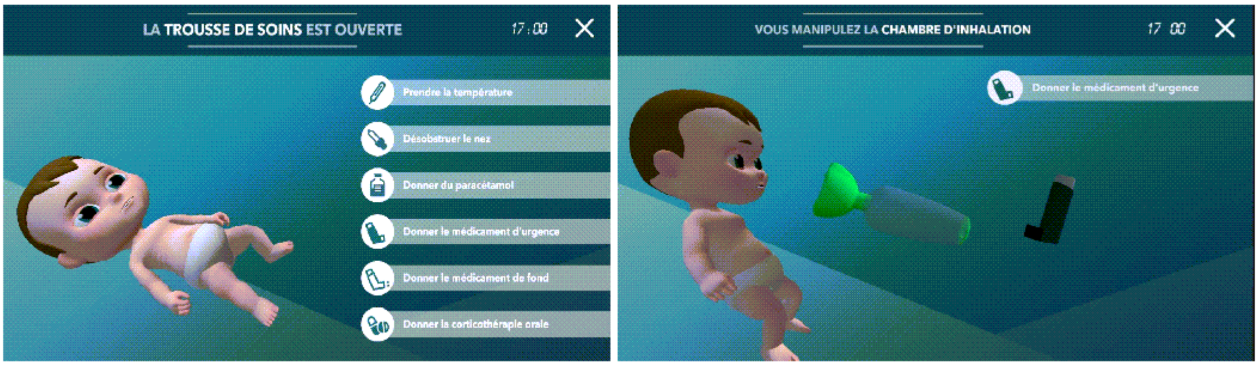
Figure 2. Simulation room with the Simbaby manikin and equipment available to the student: (1) paracetamol, (2) emergency treatment (salbutamol inhaler), (3) controller treatment (fluticasone inhaler), (4) water, (5) telephone to call 911 or pediatric emergencies, (6) oral corticosteroids (prednisolone 20 mg tablets), (7) otoscope, (8) thermometer, (9) stethoscope, (10) inhalation chamber with face mask, (11) saline pipettes, and (12) paper tissue.



The second modality used was the SG *Effic'Asthme*, a mobile app displayed on a tablet. This SG was developed by iLumens at Paris Descartes University and Dowino to train parents to recognize and manage different scenarios of asthma exacerbation in their children [16]. When connecting to *Effic'Asthme* for the first time, the user is asked to enter information to create a child avatar. The home screen then proposes the following sections: asthma action plan, asthma attack log, training, documents, and doctor (Figure 3). The training section corresponds to the SG. After completing a tutorial, the player can choose among 6 scenarios of asthma exacerbations with varying levels of severity, all taking place in the virtual bedroom of the child. A video presenting the main features of *Effic'Asthme* is available (Multimedia Appendix 2). Each scenario starts with a short briefing of the situation. A message then invites the user to carefully observe the child's avatar with the possibility of zooming and rotating the child in 3D to detect any sign of respiratory distress and to listen for a cough or wheezing. Depending on the child's condition, the user needs to choose the actions to be performed to manage the asthma exacerbation appropriately among different panels (Figure 4 left). One of the main outcomes is to check the administration technique of the emergency treatment (salbutamol/albuterol) to the child (Figure 4 right). The user can use his/her fingers to remove the cap of the spray, shake it, insert it into the spacer, place the spacer on the child's face, press the spray to administer one puff, and wait for the number of breaths they deem necessary between each puff. Once the scenario is completed, an automated, point-by-point debriefing is provided. Points are awarded for actions performed correctly, and an overall success rate of the mission out of 100% is given.

Figure 3. *Effic'Asthme* home screen: (A) "asthma action plan" automatically created by the app based on data (age, weight) entered by parent, (B) "asthma attack log" for monitoring, (C) "training" section, which allows the player to access different scenarios of simulated asthma exacerbation (only this section was used in this study), (D) "my performances," providing scores on the scenarios already done, (E) "documents," with access to detailed information on the different asthma symptoms and several questions/answers about the child's asthma, and (F) "physician" section, which helps the treating physician to know which parts of the action plan or inhalation technique should be reviewed with the family.

Figure 4. *Effic'Asthme* menu and inhalation technique page: (left) drop-down menu "emergency treatment kit" with several options (take temperature, clean nasal airway, administer paracetamol, administer emergency treatment, administer controller treatment, administer oral corticosteroids) and (right) administration of the inhaler treatment.



For the study, *Effic'Asthme* was diverted from its initial use to assess students' clinical competence. This SG was chosen because it allowed the evaluation of several dimensions of clinical competence such as knowledge and clinical reasoning and technical skills such as the administration technique of the inhaled emergency treatment.

The third assessment modality was an MCQ designed for the study (Table S2 in Multimedia Appendix 1). As with the HFS and SG, the MCQ started with a short briefing followed by 15 questions on the management of the same scenario of moderate asthma exacerbation.

Data Collection and Outcomes

For each evaluation modality, the investigators completed a standardized checklist (common for the three modalities, Table S3 in Multimedia Appendix 1). This checklist had been previously validated [16,17]. The checklist included 19 items. Each item was rated 0 or 1, depending on whether or not the student had performed the correct action when evaluated on the HFS or SG or provided the correct answer on the MCQ, with a maximum score of 19.

The main outcome was the degree of correlation between the actions performed for each item between the SG and HFS on the one hand and the concordance of the answers on the MCQ and the actions performed on the HFS on the other hand. For each item rated 0 or 1 for the modalities SG and HFS or MCQ and HFS, the Cohen kappa coefficient was calculated to estimate the concordance of the actions performed and answers provided at the level of the population studied. The median kappa coefficients on the 19 items were calculated and compared between the SG/HFS group and the MCQ/HFS group.

Secondary outcomes included the comparison between the median checklist scores for the HFS, SG, and MCQ modalities; comparison of the dispersion of the checklist scores for each modality; and participant satisfaction with the three assessment modalities.

Statistical Analysis

The sample size calculation was based on the following assumptions: 2-tailed alpha of .05, power of 80%, a median kappa coefficient estimated at 0.4 for the MCQ/HFS comparisons, and an expected median kappa coefficient of 0.7 for the SG/HFS comparisons. The calculation suggested that at least 39 participants should be enrolled [18].

Data were collected in Excel (Microsoft Corp) and analyzed with Prism software version 5.03 (GraphPad). Median values and interquartile ranges were reported. Cohen kappa coefficients for each item of the checklist between the HFS/SG and HFS/MCQ groups were calculated. The median kappa coefficients for each group were compared using a Mann-Whitney *U* test to compare the correlation of actions and answers between these two groups. The Spearman rank correlation coefficient was used to compare the median checklist scores between the three modalities. Finally, the dispersion of scores between the SG and MCQ modalities was studied by comparing variances using a modified Levene test, and participant satisfaction for each modality was compared using a student *t* test.

Results

A total of 42 students were included in the study in March and June 2019. Their characteristics are presented in Table 1. No participant was lost to follow-up.

Table 1. Participant characteristics.

Characteristics	Value (n=42)
General	
Female, n (%)	28 (67)
Age in years, median (interquartiles 1,3)	24 (23-24)
Asthma experience	
Personal history of asthma, n (%)	
Current asthma	1 (2)
Asthma during childhood	4 (10)
No history of asthma	37 (88)
Internship in general pediatrics, pediatric emergency, and/or pediatric pneumology over the past 24 months, n (%)	14 (33)
Ever witnessed an asthma attack, n (%)	7 (17)
Simulation experience, n (%)	
High-fidelity simulation experience	41 (98)
Experience with the Simbaby high-fidelity manikin	11 (26)
Video games experience, n (%)	
Ever played video games	37 (88)
Played video games in the past 6 months	15 (36)
Frequency of play, n (%)	
Never	11 (26)
Less than once a month	24 (57)
Several times a month	6 (14)
Several times a week	0 (0)
Every day or almost every day	1 (3)
Equipment owned, n (%)	
Smartphone	42 (100)
Computer	42 (100)
Digital tablet	16 (38)
Game console	18 (43)

The concordance of the actions performed during the SG session and the HFS session was moderate [19], with a median Cohen kappa coefficient (interquartile [IQ] 1-3) of 0.59 (IQ 0.45-0.69; Figure 5). The concordance of the answers provided in the MCQ and the actions performed during the HFS session was weak, with a median Cohen kappa coefficient of 0.37 (IQ 0.20-0.56). The SG offered a higher level of concordance with the actions performed during the HFS session than the MCQ as attested by the comparison of the median Cohen kappa coefficient (*P*=.04).

The median checklist scores were 9 (IQ 6-12), 8.5 (IQ 6-12), and 11 (IQ 9-14) for the HFS, SG, and MCQ modalities, respectively (Figure 6). These scores were significantly different (*P*=.01), and multiple comparisons found a significantly higher score in the MCQ group than in the HFS (*P*<.05) and SG (*P*<.05) groups.

Figure 5. Concordance of actions between the modalities assessed by Cohen kappa coefficient. Each point represents the median degree of agreement (Cohen kappa coefficient) of all students for one of the 19 items on the checklist. The horizontal bar corresponds to the median kappa coefficient for all items. HFS: high-fidelity simulation; SG: simulation game; MCQ: multiple choice questionnaire; **P*=.04.

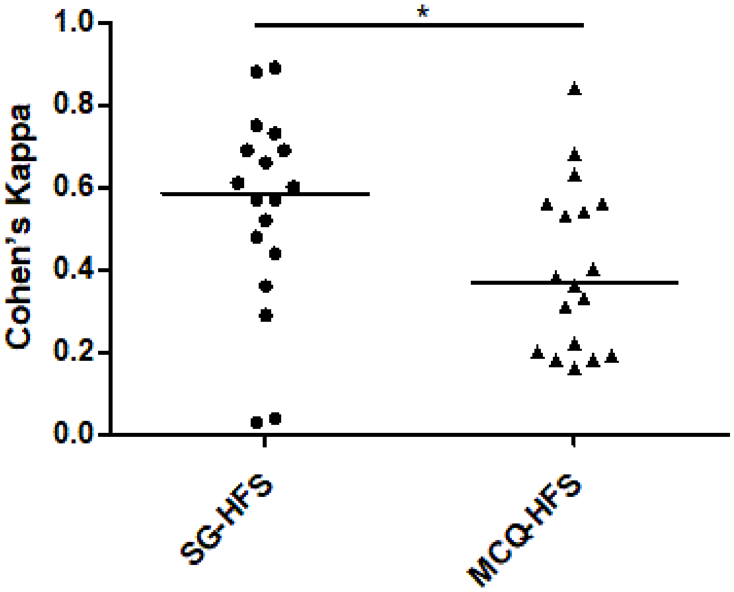
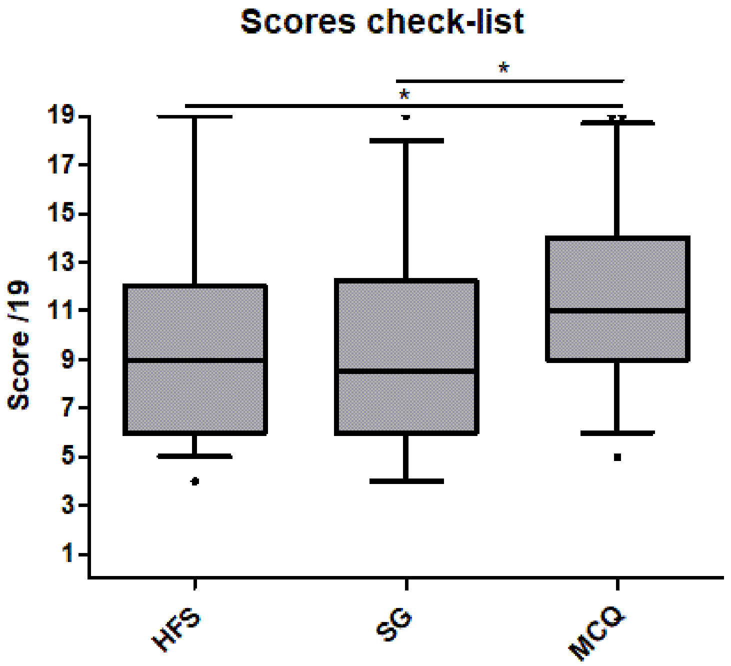


Figure 6. Median checklist scores for the modalities. HFS: high-fidelity simulation; SG: simulation game; MCQ: multiple choice questionnaire; **P*=.04.



The correlations between the SG and HFS modalities and the MCQ and HFS modalities were studied using their total checklist scores: the Spearman coefficient was 0.79 (95% CI 0.63-0.88) between the SG/HFS modalities and 0.70 (95% CI 0.50-0.83) between the MCQ/HFS modalities. This was not significantly different (*P*=.11).

The dispersion of the scores measured by the variance (σ2) of the checklist scores (17.1 and 10.9 for SG and MCQ, respectively) was not statistically significant.

Students preferred the SG modality over the MCQ (Table 2). In their opinion, the SG was more effective for progressing in the management of a clinical situation, more representative of their clinical skills, closer to real life, and more fun than MCQ (Table 2). Two-thirds of the students stated that they would prefer to be assessed in the future using SG rather than MCQ.

Table 2. Students' views of the simulation game as an educational tool in medical school.

Questions	Value, n (%)	Value, mean (SD)	*P* value
I liked this evaluation modality[a]	—[b]	—	<.001
SG[c]	—	4.5 (0.6)	—
MCQ[d]	—	3.7 (1.1)	—
I felt stress during the evaluation[a]	—	—	.26
SG	—	2.6 (1.4)	—
MCQ	—	2.2 (1.6)	—
I had fun during the evaluation[a]	—	—	<.001
SG	—	3.9 (0.8)	—
MCQ	—	1.9 (1.5)	—
I would say that the most effective modality to progress in the management of a clinical situation such as a child asthma exacerbation is...			
SG	28 (67)	—	—
MCQ	14 (33)	—	—
I would say that the modality most representative of reality is...			
SG	41 (98)	—	—
MCQ	1 (2)	—	—
I would say that the modality most representative of my skills is...			
SG	38 (90)	—	—
MCQ	4 (10)	—	—
I would personally prefer to be evaluated in the future by...			
SG	27 (64)	—	—
MCQ	15 (36)	—	—

[a] Results are expressed as a mean (SD) between 1 (strongly disagree) and 5 (strongly agree).

[b] Not applicable.

[c] SG: simulation game.

[d] MCQ: multiple choice questionnaire.

Discussion

Principal Findings

HFS and OSCE, the best available modalities for assessing the clinical competence of medical students without any risk to the patient, are almost impossible to implement on a large scale due to their high consumption of time and human resources. Thus, many faculties use MCQ as a default solution for large-scale examinations, with the risk of favoring students who are highly competent to respond to MCQs but unable to manage real-life emergency situations. This study demonstrates that *Effic'Asthme*, an SG created to teach about the management of pediatric asthma exacerbations, better reflects the actions of fifth-year medical students on HFS than an MCQ on the same clinical case. We propose in Table 3 a summary of the domains of clinical competence that can be assessed through an SG such as *Effic'Asthme*, in comparison with MCQ and HFS. As illustrated, if further studies confirm these findings, SG may become an intermediate solution between HFS and MCQ, evaluating students' clinical competence further than MCQ while remaining widely deployable, unlike HFS.

Table 3. Modalities of assessment and domains covered.

Domain	Multiple choice questionnaires	Simulation game	High-fidelity simulation
Knowledge (eg, corticosteroids dose)	+++[a]	+++ Possible with *Effic'Asthme*	+++
Clinical reasoning (eg, choosing between repeating salbutamol or calling emergency medical services after 20 minutes)	++	+++ Possible with *Effic'Asthme*	+++
Technical skills (eg, inhalation technique when using a spacer)	–[b]	+ Possible with *Effic'Asthme*	+++
Recognition of symptoms (eg, chest-indrawing, audible wheezing)	–	+++ Possible with *Effic'Asthme*	+++
Communication skills (eg, reassuring the parent)	–	– Not possible with *Effic'Asthme*	++
Teamwork (eg, coordinating with a nurse)	–	++ (multiplayer) Not possible with *Effic'Asthme*	+++

[a] Appropriate to evaluate the domain.

[b] Not appropriate to evaluate the domain.

Contextualizing our results within the medical literature is limited by the paucity of data on the comparison of SG with other assessment modalities. To our knowledge, only the study by Adjedj et al [15] compared the SG and MCQ modalities for the evaluation of medical students, on a clinical case of cardiology. They found greater variability in the scores obtained on the SG than in the scores obtained on the MCQ and concluded that SG had the potential to better rank students [15]. However, because their study did not use a gold standard as ours did, they were unable to determine which of the two assessment modalities was more relevant for evaluating students. Therefore, our study goes further in showing that SG is indeed a more relevant assessment modality than MCQ for the evaluation of the management of a child's asthma exacerbation, as it better reflects the actions performed on HFS. Surprisingly, whereas the analysis, item by item, of the actions performed/answers provided revealed a higher level of concordance between the SG/HFS modalities than between the MCQ/HFS modalities, this difference was no longer statistically significant when the total checklist scores were compared (P=.11). The most likely explanation is a lack of power because our sample size was not calculated on this outcome.

After considering the SG *Effic'Asthme* as a new evaluation modality, its acceptability and cost should be scrutinized. The results of the survey conducted with the medical students who participated in this study indicate that students support the use of this evaluation modality for their exams, as reported in other studies [20,21]. It would be interesting to collect the opinions of the members of the faculty to see if they are consistent with those of students. Clearly, faculty members will need to estimate the costs of this new assessment modality. Developing SG is costly: their development is costly. For example, the development of *Effic'Asthme* cost €135,000 (US $160,000) [22]. This cost could be reduced in the future by using an interuniversity platform integrating all the necessary elements to create a multitude of scenarios in the same way that high-fidelity simulators are provided with a software allowing to design hundreds of different situations.

A limitation of the SG revealed by this study was the need to integrate a tutorial. Indeed, in our study, many participants did not perform actions such as shaking the inhaler or removing its cap, although they had performed these actions on the HFS and the MCQ. This suggests that they thought it was impossible to perform these actions on the SG. Thus, based on the results of this study, a key recommendation for future developers of SGs is to develop a neutral tutorial before the assessment part of their SG.

Limitations

The study has several limitations. First, it can be argued that the order of assessments (HFS then SG then MCQ) may have favored higher scores on the MCQ modality. Initially, we considered a crossover trial: after their evaluation with the HFS modality, students would have been randomized into two groups, one group starting by the SG session and ending with the MCQ modality as in this study, and the other starting by the MCQ modality and ending with the SG session. However, the MCQ required students to be provided several cues (eg, when asking for the dose of SABAs or oral corticosteroids, it was implied that SABAs and oral corticosteroids should be administered). This would have strongly biased the next evaluation on SG. By contrast, the HFS and SG modalities did not influence the student on specific actions to be taken because no indication was provided during or after the sessions, and we judged that there was no reason for students to perform better over the course of the different assessments without being provided any indication between them. Indeed, it was demonstrated that no gain in knowledge can be achieved when no debriefing is provided following a simulation session [23]. The use of HFS as a gold standard for assessing medical students' competence is also debatable. HFS is probably not a modality that is perfectly representative of the care that would have been provided by the students under real-world conditions, but it corresponded to the only option that we have found to be both feasible and ethical. Finally, the fact that the study was monocentric and evaluated only volunteer students on a single scenario limits the generalizability of the results. Despite these limitations, this trial is the first to compare SG and MCQ as assessment methods for medical students.

Conclusion

In conclusion, the SG used in this study better reflected the clinical competence of students on HFS than an MCQ on the same clinical case of pediatric asthma exacerbation. Its use as an assessment method was appreciated by the students. If further studies confirm these results, SG could become an interesting compromise for the evaluation of medical students between the cheap but limited assessment allowed by MCQs and the comprehensive but highly expensive assessment allowed by HFS.

Authors' Contributions

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

DD and AH participated to the creation of the simulation game *Effic'Asthme*.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 17 KB - jmir_v22i12e23254_app1.docx]

Multimedia Appendix 2
Example of a scenario in *Effic'Asthme*.
[MP4 File (MP4 Video), 180013 KB - jmir_v22i12e23254_app2.mp4]

References

1. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002 Jan 09;287(2):226-235. [Medline: 11779266]
2. Norcini JJ, McKinley DW. Assessment methods in medical education. Teaching Teacher Educ 2007 Apr;23(3):239-250. [doi: 10.1016/j.tate.2006.12.021]
3. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001 Mar 24;357(9260):945-949. [doi: 10.1016/S0140-6736(00)04221-5] [Medline: 11289364]
4. Epstein RM. Assessment in medical education. N Engl J Med 2007 Jan 25;356(4):387-396. [doi: 10.1056/NEJMra054784] [Medline: 17251535]
5. Doughty CB, Kessler DO, Zuckerbraun NS, Stone KP, Reid JR, Kennedy CS, et al. Simulation in pediatric emergency medicine fellowships. Pediatrics 2015 Jul;136(1):e152-e158. [doi: 10.1542/peds.2014-4158] [Medline: 26055850]
6. Alvarez J. Du Jeu Vidéo au Serious Game?: Approches Culturelle, Pragmatique et Formelle [Thesis]. Toulouse: University of Toulouse; 2007.
7. Alvarez J, Djaouti D. Une taxinomie des Serious Games dédiés au secteur de la santé?: L'application des jeux vidéo à la médecine et à la surveillance. Rev électr électron. 2008. URL: https://docplayer.fr/16061373-Une-taxinomie-des-serious-games-dedies-au-secteur-de-la-sante.html [accessed 2020-11-28]
8. Sitzmann T. A meta-analytic examination of the instructional effectiveness of computer-based simulation games. Personnel Psychol 2011;64(2):528. [doi: 10.1111/j.1744-6570.2011.01190.x]
9. Drummond D, Hadchouel A, Tesnière A. Serious games for health: three steps forwards. Adv Simul 2017 Feb 4;2(1):3. [doi: 10.1186/s41077-017-0036-3]
10. Charsky D. From edutainment to serious games: a change in the use of game characteristics. Games Culture 2010 Apr;5(2):177-198. [doi: 10.1177/1555412009354727]
11. Pasin F, Giroux H. The impact of a simulation game on operations management education. Comput Educ 2011 Aug;57(1):1240-1254. [doi: 10.1016/j.compedu.2010.12.006]
12. Salas E, Wildman JL, Piccolo RF. Using Simulation-Based Training to Enhance Management Education. AMLE 2009 Dec;8(4):559-573. [doi: 10.5465/amle.8.4.zqr559]
13. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012 Oct;99(10):1322-1330. [doi: 10.1002/bjs.8819] [Medline: 22961509]
14. Wang R, DeMaria S, Goldberg A, Katz D. A systematic review of serious games in training health care professionals. Simul Healthc 2016 Feb;11(1):41-51. [doi: 10.1097/SIH.0000000000000118] [Medline: 26536340]
15. Adjedj J, Ducrocq G, Bouleti C, Reinhart L, Fabbro E, Elbez Y, et al. Medical student evaluation with a serious game compared to multiple choice questions assessment. JMIR Serious Games 2017 May 16;5(2):e11 [FREE Full text] [doi: 10.2196/games.7033] [Medline: 28512082]
16. Drummond D, Fabbro E, Tesnière A, Hadchouel A. Toward virtual simulation for parents of children with asthma. J Allergy Clin Immunol Pract 2017;5(6):1779-1781. [doi: 10.1016/j.jaip.2017.04.017] [Medline: 28529084]

17. Drummond D, Truchot J, Fabbro E, Ceccaldi P, Plaisance P, Tesnière A, et al. Fixed versus variable practice for teaching medical students the management of pediatric asthma exacerbations using simulation. Eur J Pediatr 2018 Feb;177(2):211-219. [doi: 10.1007/s00431-017-3054-1] [Medline: 29204851]
18. Bujang M, Baharum N. Guidelines of the minimum sample size requirements for Kappa agreement test. Epidemiol Biostatistics Public Health 2017;14(2):122671-1226710. [doi: 10.2427/12267]
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
20. Bick JS, Demaria S, Kennedy JD, Schwartz AD, Weiner MM, Levine AI, et al. Comparison of expert and novice performance of a simulated transesophageal echocardiography examination. Simul Healthc 2013 Oct;8(5):329-334. [doi: 10.1097/SIH.0b013e31829068df] [Medline: 24030477]
21. Chon S, Timmermann F, Dratsch T, Schuelper N, Plum P, Berlth F, et al. Serious games in surgical medical education: a virtual emergency department as a tool for teaching clinical reasoning to medical students. JMIR Serious Games 2019 Mar 05;7(1):e13028 [FREE Full text] [doi: 10.2196/13028] [Medline: 30835239]
22. Drummond D, Tesniere A, Hadchouel A. Design and development of a simulation game for parents of preschool children with asthma: an evidence-based approach. Eur Respiratory Soc 2018;52:PA5445. [doi: 10.1183/13993003.congress-2018.pa5445]
23. Shinnick MA, Woo M, Horwich TB, Steadman R. Debriefing: the most important component in simulation? Clin Simul Nurs 2011 May;7(3):e105-e111. [doi: 10.1016/j.ecns.2010.11.005]

Abbreviations

HFS: high-fidelity simulation
IQ: interquartile
MCQ: multiple choice questionnaires
OSCE: objective structured clinical examination
SABA: short-acting beta-agonist
SG: simulation game

Edited by G Eysenbach, R Kukafka; submitted 05.08.20; peer-reviewed by Z Aghaei, A Khaleghi; comments to author 18.09.20; revised version received 10.10.20; accepted 12.11.20; published 16.12.20.

Please cite as:
Fonteneau T, Billion E, Abdoul C, Le S, Hadchouel A, Drummond D
Simulation Game Versus Multiple Choice Questionnaire to Assess the Clinical Competence of Medical Students: Prospective Sequential Trial
J Med Internet Res 2020;22(12):e23254
URL: http://www.jmir.org/2020/12/e23254/
doi:10.2196/23254
PMID:33325833

Original Paper

Perceived Impact of an Online Community Care Platform for Dutch Older Adults on Local Participation, Informal Caregiving, and Feelings of Connectedness: Pretest-Posttest Observational Study

Sarah Willard[1,2], MSc; Erik van Rossum[1,2], PhD; Marieke Spreeuwenberg[1,2], PhD; Luc de Witte[1,3], PhD

[1]Centre of Innovative Care and Technology (EIZT), Zuyd University of Applied Sciences, Heerlen, Netherlands

[2]Care and Public Health Research Institute, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[3]Centre for Assistive Technology and Connected Healthcare , University of Sheffield, Sheffield, United Kingdom

Corresponding Author:
Sarah Willard, MSc
Centre of Innovative Care and Technology (EIZT)
Zuyd University of Applied Sciences
Henri Dunantstraat 2
Heerlen, 6419PB
Netherlands
Phone: 31 618212931
Email: sarah.willard@zuyd.nl

Abstract

Background: In a changing ageing society wherein older adults are increasingly expected to take care of themselves instead of relying on health care services, online community care platforms can help older adults to meet these expectations. A considerable number of these online community care platforms have been introduced in several European countries based on their potential. However, their actual impact is unclear.

Objective: The aim of this study was to investigate the self-reported use, expectations, and perceived impact of a Dutch online community care platform called Grubbenvorst-Online among Dutch older adults. The following 2 questions were studied: (1) What is the self-reported use of Grubbenvorst-Online among older adults? (2) What are their expectations and perceived impact of Grubbenvorst-Online regarding local participation, their social network, mutual informal caregiving, and feelings of connectedness?

Methods: An observational pretest-posttest study was conducted. Participants were recruited via a web-based message on the Grubbenvorst-Online platform and data were collected via postal questionnaires among older users at the start of the study and 4 months later. Data regarding the expectations and the perceived impact of Grubbenvorst-Online were compared and tested.

Results: Forty-seven Grubbenvorst-Online users with an average age of 74 years participated in this study. They were healthy, predominantly "internet-skilled," and they found the internet important for maintaining social contacts. In general, the use of the online community care platform decreased during the 4-month follow-up period. The perceived impact of Grubbenvorst-Online was significantly lower than that expected regarding information provision (P=.003), seeking help from fellow villagers (P<.001), giving help to fellow villagers (P<.001), and consulting care or welfare services (P<.001).

Conclusions: The findings of this study indicate that online community care platforms perhaps do not provide enough "added value" in their current form. We suggest a new direction in which online community care platforms primarily support existing offline initiatives aimed at stimulating local participation, informal caregiving, and feelings of connectedness.

(J Med Internet Res 2020;22(12):e20304) doi:10.2196/20304

KEYWORDS

older adults; online community; online platform; social network; local participation; informal caregiving

Introduction

In a changing ageing society wherein older adults are increasingly expected to take care of themselves instead of relying on health care services [1,2], online community care platforms can help older adults to meet these expectations. A considerable number of web-based platforms for older adults has been introduced in several European countries, for example, Germany [3], the United Kingdom [4], Belgium, and the Netherlands [5-8]. All these platforms target older adults and their care networks to facilitate and support aging-in-place.

In our scoping review, a typology of online care platforms for community-dwelling older adults was developed. This review was performed because little research had been conducted on the availability of web-based platforms for older adults and their characteristics, functionalities, and usability in order to guide older adults in choosing a suitable platform. The review resulted in an overview of 21 care platforms, which can be classified into the following 3 types: (1) *Online Community Care Platforms*, which attempt to enhance social cohesion by interlinking community-dwelling older adults with neighboring informal caregivers and by promoting local activities at the neighborhood level; (2) *Online Care Network Platforms*, which provide older adults and professional and informal caregivers tools to coordinate, plan, and communicate about (health) care; and (3) *System Integrator Platforms*, which interconnect a variety of functionalities. The latter platform type has the capability of integrating existing services and apps into its own software, that is, it operates as an "empty" information communication technology framework, which can be filled with any content [9]. This study focuses on a *System Integrator Platform* that was deployed as an online community care platform. In other words, all functionalities in this information communication technology framework were aimed at enhancing social cohesion or at promoting local activities at the neighborhood level. Thus, we choose to use the term "online community care platform" when referring to the platform in this study.

Online community care platforms offer older adults various apps aimed at supporting their independent living (eg, products and services) and civic and social participation (eg, contacts, messages, a matching tool for informal care). Studies have shown that more common web-based communities for older adults, which are known under different names such as "social networking sites" (eg, Facebook, Twitter, LinkedIn), "online social networks," or "online social communities" can potentially have a positive impact on civic participation and help to develop and maintain social relationships and social support [10-13].

It seems that online community care platforms are implemented in several European countries based on the assumption that they can support older adults to age-in-place, to participate locally, and that they can help to develop and maintain social relationships and to arrange social support. As governments call for increased autonomy at local levels [1], these platforms seem perfectly suited tools for older adults to actually remain or even become more autonomous. However, their *actual* (perceived) impact is unclear [9]. The majority of previous research has focused on the usability and acceptability of these web-based communities or platforms and thus on the preimplementation phase. Hardly any studies report the factors that contribute to the continued use or to the desertion of web-based communities once they have been implemented [14]. It is important for both users and policy makers to discover to what extent these platforms actually help older adults to participate locally and socially.

This study therefore primarily focusses on the perceived impact in the postimplementation phase of an online community care platform that was implemented in 2015, called as "Grubbenvorst-Online" (abbreviated as GO, Grubbenvorst refers to the town in which the platform was implemented). Grubbenvorst-Online was the initiative of an active group of local older citizens ("the initiators") who, through the platform, aimed to help both older adults and vulnerable inhabitants of Grubbenvorst to socially participate locally. Grubbenvorst-Online was founded by the initiators in close collaboration with local entrepreneurs, associations, and social and health care organizations. Arestoco, a Belgian enterprise, provided the information communication technology framework for the Grubbenvorst-Online platform. Their platform entitled "Cubigo" was selected, as its software could be modified based on the wishes and needs of potential users [15]. The content of Grubbenvorst-Online (Figure 1) was determined and kept up-to-date by the initiators.

A few examples of the platform's apps are (1) a matching tool for informal care, called as "Help each other," in which users can exchange informal help (Figure 2), (2) a local calendar where information can be found on local events and activities (Figure 3), and (c) social services in which users can find information about available care services and organizations.

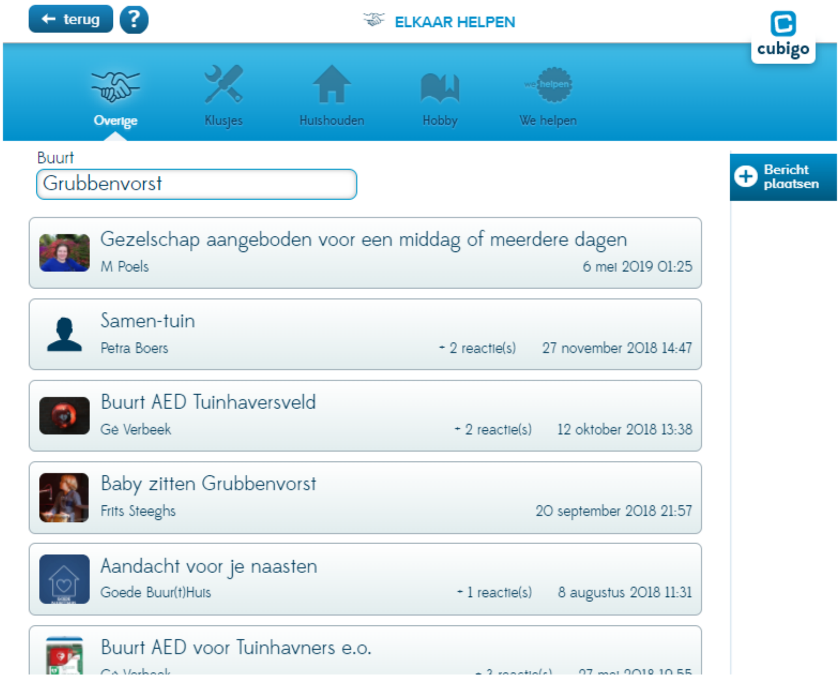
The following 2 questions were studied.

1. What is the self-reported use of Grubbenvorst-Online among older adults?
2. What are the expectations and perceived impact of Grubbenvorst-Online among older adults regarding local participation, their social network, mutual informal caregiving, and feelings of connectedness?

We wanted to determine what users expected from Grubbenvorst-Online when they started using it, what did they think the platform was intended for and had to offer in general, and then to investigate to what extent the platform met these expectations.

Figure 1. Screenshot of the online community care platform Grubbenvorst-Online.

Figure 2. Screenshot of the functionality "Help each other".



Grubbenvorst

Gezelschap aangeboden voor een middag of meerdere dagen
M Poels 6 mei 2019 01:25

Samen-tuin
Petra Boers - 2 reactie(s) 27 november 2018 14:47

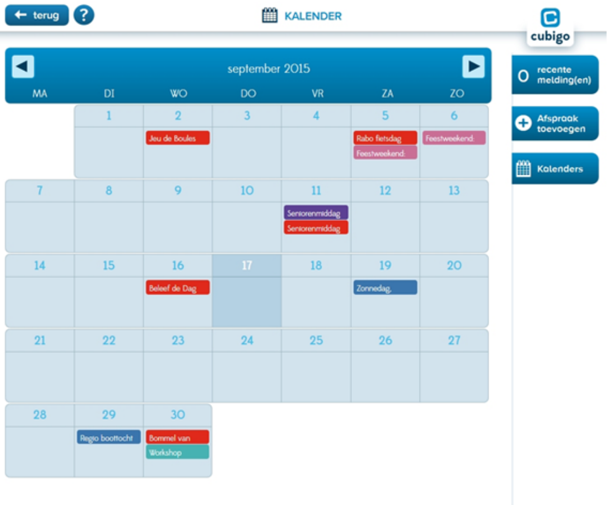
Buurt AED Tuinhaversveld
Gé Verbeek - 2 reactie(s) 12 oktober 2018 13:38

Baby zitten Grubbenvorst
Frits Steeghs 20 september 2018 21:57

Aandacht voor je naasten
Goede Buur(t)Huis - 1 reactie(s) 8 augustus 2018 11:31

Buurt AED voor Tuinhavners e.o.

Figure 3. Screenshot of the functionality "Calendar".



Methods

Design

This study had an observational pretest-posttest design. Data on the use, expectations, and perceived impact regarding the online community care platform Grubbenvorst-Online was collected via 2 postal questionnaires completed by older users. Our aim was to perform an exploratory study among approximately 50 participants.

Setting

Grubbenvorst is a sparsely populated village in Limburg, which is the southernmost province of the Netherlands. Approximately 1200 of the 4800 inhabitants (25.0%) are 65 years or older [16]. Grubbenvorst-Online was made available for all inhabitants of Grubbenvorst in August 2015. Several activities were undertaken to communicate and promote the existence and availability of the platform, such as (1) public launch of Grubbenvorst-Online in the central square in Grubbenvorst in the presence of the alderman; (2) publication of advertisements in “Announcements” (a local newspaper); (3) distribution of flyers, brochures, and other public relations material; (4) placement of a large billboard at the main entrance of the village; (5) various informative presentations for organizations such as associations for older adults, care organizations, and an elementary school; and (6) taking part in a TV interview at the local broadcaster (TV Reindonk). At the start of the study in early 2018, Grubbenvorst-Online had approximately 725 users.

Participants

Participants were recruited among the users in January 2018 via a web-based message on the Grubbenvorst-Online platform. This message described the purpose of the study, that is, to gather information regarding the way in which older adults used the platform and how they perceived its impact. Older adults were asked, if interested, to register within 1 month by leaving their contact details with the initiators. Applicants were suitable for inclusion if they were 65 years or older and if they were a user of the online community care platform. This means that every potential participant was an existing user, had already registered on the online community care platform, and was therefore more or less familiar with the functionalities and operation of the platform. Furthermore, the length of time that the participants used the platform was *not* an inclusion criterion; hence, both “old” and “new” users were suitable for inclusion. Recruitment stopped by the time 51 participants had registered.

Statement of Ethical Approval

The ethical principles that are outlined in the Dutch “Medical Research Involving Human Subjects Act” were followed throughout the entire study. The ethical approval for this study was given by METC-Z (16-N-213).

Data Collection

Two postal questionnaires were used to collect reports of older adults regarding the use, expectations, and perceived impact of the Grubbenvorst-Online platform. Potential participants received an invitation letter, informed consent form, and the first (baseline) postal questionnaire (T0) in March 2018. Four months later, in July 2018, they received the second postal questionnaire (T1). Table 1 illustrates the themes and topics of the questionnaire and their operationalization. The baseline questionnaire (T0) was primarily designed to collect information on the *expectations* of older adults regarding the online community care platform: did participants expect the platform to have positive effects on local participation, informal caregiving, their social network, and feelings of connectedness? The follow-up questionnaire (T1) aimed to discover to what extent the expectations as formulated in T0 were met (ie, perceived impact). As there were no validated scales available to measure the expectations and perceived impact of online community care platforms, the scale was developed by us.

General characteristics of the participants were collected as well as details of their use of the internet and technology and use of informal care and their community involvement and related needs. Regarding the aforementioned characteristics, validated scales or parts of existing questionnaires were used. For topics 2-4 (Table 1), parts of the validated questionnaire “Senior Monitor Heerlen” [17] were used, that is, “Civic participation and social network” and “Living and environment.” Additionally, for topic 4, the validated scale “Involvement with neighbors” [18] was adopted. This scale incorporates 3 domains of social cohesion: (1) collaboration to stimulate local well-being; (2) solidarity: the extent to which neighbors help each other, and (3) feelings of connectedness: the extent to which neighbors feel connected to each other.

Concerning the topic of informal care, participants could answer yes or no to whether they gave help, whether they were willing to give help, whether they already received help, and finally, whether they would accept help from someone living in their close vicinity. If participants indicated that they helped someone, they did not have to answer the question whether they were prepared to offer help. The topic was introduced as follows: “the following questions are about your ability to help others or to receive help yourself. This “help” can, for example, consist of doing chores around the house or doing groceries together with family, friends, neighbors, or acquaintances. Thus, the questions were not related to physical care tasks.

Table 1. Questionnaire themes and their operationalization.

Theme	T0/T1[a]	Items (n)	Topics	Adopted propositions and questions
Participant characteristics	T0	27	(1) General characteristics (eg, gender, date of birth), (2) Use of internet and technology, (3) Informal care, (4) Community involvement and related needs	Examples of questions for topics 2-4: (2) How important is the internet for you to keep in touch with other people? (nominal 5-point scale from “very important” to “unimportant”), (3) Would you accept help from a neighbor? (Yes/No), (4) Do you feel involved with the people living in your immediate vicinity? (nominal 5-point scale: “With almost none,” “With most not,” “With some,” “With most,” “With almost everyone”).
Self-reported use of Grubbenvorst-Online	T0-T1	14-17[b]	Use of platform in general and of specific functionalities[c]	Examples of questions (nominal 4-point scale: “Never,” “Occasionally,” “Regularly,” and “Daily”): (1) How often do you currently use the Grubbenvorst-Online platform? (T0 + T1); (2) How often do you currently use the “Grubbenvorst Village Calendar”? (T0+T1); (3) Have you used Grubbenvorst-Online less, more often, or to the same extent compared to 4 months ago? (T1)
Expectations of Grubbenvorst-Online	T0	7	Expectations regarding the platform’s added value regarding local participation, informal caregiving, social networking, and feelings of connectedness	Examples of propositions (3-point scale: “[Strongly] Agree,” “Neither agree nor disagree,” “[Strongly] Disagree”): (1) I expect I will partake more in village activities because of Grubbenvorst-Online, (2) I expect that I will use Grubbenvorst-Online to ask my fellow villagers for help (eg, with a job around the house or grocery shopping)
Perceived impact of Grubbenvorst-Online	T1	7	Perceived impact of the platform regarding local participation, informal caregiving, social network, and feelings of connectedness	Examples of propositions (3-point scale: “[Strongly] Agree,” “Neither agree nor disagree,” “[Strongly] Disagree”): (1) Because of Grubbenvorst-Online, I partake more often in village activities. (2) I asked for the help of a fellow villagers via Grubbenvorst-Online (eg, for a job in the house or grocery shopping)

[a] T0: baseline questionnaire, T1: questionnaire after 4 months of the study.

[b] Apart from a few additional questions in T1 (such as example 3), the questions were identical in T0 and T1.

[c] Not all functionalities of Grubbenvorst-Online (see Figure 1) were assessed. Only functionalities identified by the initiators as most important were included in the questionnaire.

Data Analysis

The baseline characteristics of the participants who filled out both questionnaires (T0 and T1) were expressed in mean (SD) or n (% of the participants). The comparison of the expectations and perceived impact of Grubbenvorst-Online was tested with the McNemar test. All *P* values were two-sided and were considered to be statistically significant if less than .05. SPSS 25 (IBM SPSS Statistics for Windows, IBM Corp) was used for data entry and statistical analysis.

Results

Participant Characteristics

Of the 51 initial participants, 47 were included in the data analysis as they filled out both questionnaires (T0 and T1). This group consisted of 25 men and 22 women with a mean (SD) age of 74 (6.2) years. All participants were of Dutch nationality and had a relatively high level of education: 18 out of 47 (38%) had a university degree or a higher professional education qualification and 19 participants (40%) had a secondary vocational education qualification. The majority of the participants (32/47, 68%) lived with a partner, either married or unmarried. The average grade that participants gave for their own health was 7.6 (range 0-10, higher scores indicating better health). Only 14 participants (30%) indicated that they felt slightly hindered in their daily activities due to a long-term illness or disability.

Use of Technology and Internet

All 47 participants were asked to indicate the extent to which they adopted certain devices. A smartphone was used "regularly" by 11 participants (23%) and "daily" by 18 participants (38%). An iPad or tablet was used "regularly" or "daily" by 26 participants (55%), and a laptop was used "daily" by 26 participants (55%). All participants indicated that they used the internet. Furthermore, 31 out of the 47 participants (66%) considered its use to be "(very) easy," 12 (25%) participants found the easiness of internet use as neutral, and only 4 participants (9%) found the use of internet "difficult." In addition, 30 participants (65%) considered the internet to be "(very) important for keeping in touch with others."

Informal Care

Out of the 47 participants, 13 (28%) indicated that they helped someone living in their immediate vicinity. A considerably larger proportion (31/47) was prepared to offer help and the majority (43/47) was willing to accept help. However, 39 (83%) participants indicated that they did not need any help from people living in their immediate vicinity (see Table 2).

Table 2. Participants' views on informal care (n=47).

Variables	Value
Gave help to someone, n (%)	
Yes	13 (28)
No	33 (70)
Missing (no answer)	1 (2)
Prepared to offer help, n (%)	
Yes	31 (66)
No	6 (13)
Missing (no answer)	10 (21)
Willing to accept help, n (%)	
Yes	43 (92)
No	3 (6)
Missing (no answer)	1 (2)
Need help, n (%)	
Yes	7 (15)
No	39 (83)
Missing (no answer)	1 (2)

Community Involvement and Related Needs

The extent to which participants felt connected to people in their immediate vicinity was high; 20 out of 47 participants (43%) felt connected with most people and 13 participants (28%) with almost everyone. Only 5 participants (11%) indicated a need for more contact with people from their immediate vicinity. The vast majority, that is, 33 out of 47 participants (70%) specified that they felt no need for more contacts as they were satisfied with the number of contacts they already had. These results show that, in general, the participants already had a relatively large social network and felt no need to expand it further.

Use of the Grubbenvorst-Online Community Care Platform

The mean (SD) score (range 0-10, higher scores indicating greater valuation) that the participants gave to the online community care platform as a whole was 7.3 (1.0) at T0 and 7.1 (1.3) at T1. Table 3 gives an overview of the use of the Grubbenvorst-Online platform as a whole (see row and per functionality at baseline, T0, and at follow-up, T1). The functionalities "Grubbenvorst-Online," "Columns and interviews," "Photo and Video," and "Messages" were the most frequently used at both T0 and T1. However, only a minority of the participants indicated that they used the aforementioned functionalities on a regular or daily basis. The functionalities "Dialect Dictionary," "Social Services," "Local entrepreneurs," "Local associations," "Health care services," and "Contacts" were the least used.

In general, a minor decrease was reported in the use at the level of specific functionalities: at T1, the "regular" use of functionalities decreased or remained stable while the "daily" use of functionalities did not occur (with the exception of 2 participants who consulted the Grubbenvorst-Online feature).

Table 3. Self-reported use of the online community care platform Grubbenvorst-Online as a whole and per functionality (n=47).

Online community care platform/functionality	Regular use	Daily use
Grubbenvorst-Online (platform), n (%)		
T0[a]	17 (36)	4 (9)
T1[b]	11 (23)	3 (6)
Grubbenvorst-Online (functionality), n (%)		
T0	14 (30)	1 (2)
T1	12 (26)	2 (4)
Columns and interviews, n (%)		
T0	12 (26)	0 (0)
T1	7 (15)	0 (0)
Calendar Grubbenvorst, n (%)		
T0	8 (17)	0 (0)
T1	8 (17)	0 (0)
Photo and video, n (%)		
T0	11 (23)	0 (0)
T1	11 (23)	0 (0)
Dialect dictionary, n (%)		
T0	3 (6)	0 (0)
T1	2 (4)	0 (0)
Social services, n (%)		
T0	2 (4)	0 (0)
T1	2 (4)	0 (0)
Local entrepreneurs, n (%)		
T0	2 (4)	0 (0)
T1	0 (0)	0 (0)
Local associations, n (%)		
T0	4 (9)	0 (0)
T1	4 (9)	0 (0)
Health care services, n (%)		
T0	4 (9)	0 (0)
T1	2 (4)	0 (0)
Contacts, n (%)		
T0	6 (13)	0 (0)
T1	2 (4)	0 (0)
Messages, n (%)		
T0	15 (33)	2 (4)
T1	12 (26)	0 (0)
Help each other, n (%)		
T0	6 (13)	0 (0)
T1	6 (13)	0 (0)

[a] T0: baseline questionnaire.

[b] T1: questionnaire after 4 months of the study.

Impact of the Grubbenvorst-Online Platform: Expectations and Experiences

Table 4 shows the number of participants who (fully) agreed at T0 with propositions about various potential effects of their use of Grubbenvorst-Online; in other words, these correspond to the participants' expectations of the platform. Table 4 also shows the number of participants who (fully) agreed with the same propositions about Grubbenvorst-Online at T1: this is the participants' perception of the impact of the platform after having used it. In general, the participants' expectations of Grubbenvorst-Online were not fully met. At T1, the participants' overall perceived impact of Grubbenvorst-Online was significantly lower than they had expected with respect to "information provision about Grubbenvorst" (*P*=.003), "seeking help from fellow villagers" (*P*<.001), "giving help to fellow villagers" (*P*<.001), and "consulting care or welfare services" (*P*<.001). Their expectations of Grubbenvorst-Online and its perceived impact differed least regarding "participating in local activities," "feeling connected to Grubbenvorst," and "expansion of social network." Overall, participants perceived the highest impact of Grubbenvorst-Online regarding "information provision about Grubbenvorst," "feeling connected to Grubbenvorst," and "participating in local activities."

Table 4. Expectations of Grubbenvorst-Online (T0) and the perceived impact of Grubbenvorst-Online (T1) regarding various indicators.

Variables	Value (n=47)	*P* value
Information provision about Grubbenvorst, n (%[a])		.003[b]
Expectation	36 (77)	
Perceived impact	24 (51)	
Participating in local activities, n (%)		.23
Expectation	14 (30)	
Perceived impact	9 (19)	
Feeling connected to Grubbenvorst, n (%)		.18
Expectation	19 (40)	
Perceived impact	14 (30)	
Expansion of social network, n (%)		.11
Expectation	13 (28)	
Perceived impact	6 (13)	
Seeking help from fellow villagers, n (%)		<.001[b]
Expectation	17 (36)	
Perceived impact	2 (4)	
Giving help to fellow villagers, n (%)		<.001[b]
Expectation	17 (36)	
Perceived impact	3 (6)	
Consulting care or welfare services, n (%)		<.001[b]
Expectation	22 (47)	
Perceived impact	1 (2)	

[a] Strongly agree with proposition.

[b] This value was significant at *P*<.05 in the McNemar test (two-sided).

Discussion

Major Findings

We explored the self-reported use, expectations, and perceived impact among older adults of a Dutch online community care platform. The study involved 47 healthy and predominantly "internet-skilled" older users (average age, 74 years). The vast majority of these users indicated that they were willing to help people living in their immediate vicinity; however, they did not necessarily feel a "help need" themselves. Furthermore, only a small proportion indicated a need for more social contact. The online community care platform was graded by the participants with an overall "more than sufficient" grade of 7.2. In general, the use of the online community care platform decreased during the 4-month follow-up period. The functionalities "Grubbenvorst-Online," "Messages," "Photo and Video," and "Columns and interviews" were the most frequently used. At follow-up, participants' perceived impact of Grubbenvorst-Online was significantly lower than their initial expectations of the impact the platform would provide regarding "information provision about Grubbenvorst," "seeking help from fellow villagers," "giving help to fellow villagers," and "consulting care or welfare services."

When the participants' perceived impact of Grubbenvorst-Online is viewed in isolation (ie, without comparing it to their expectations), it can be ascertained that participants perceived the highest impact with respect to "information provision about Grubbenvorst," "feeling connected to Grubbenvorst," and "participating in local activities." This indicates that an online community care platform can play a role in stimulating positive feelings toward communities and encouraging its members to undertake activities locally.

It is noteworthy that participants perceived little-to-no impact regarding (mutual) informal caregiving (ie, asking for or giving help to fellow villagers), although there was a high degree of willingness on the part of these participants to care for or accept help from people living in their vicinity. However, the participating older adults did indicate that they felt no need to expand their networks and did not express any great need for care from people living in their vicinity. Therefore, perhaps this limited perceived impact regarding (mutual) informal caregiving is related to the limited need for support or network expansion. This finding is remarkable as these web-based communities or platforms aim to support (and stimulate) older adults to exchange various forms of social support as it is increasingly expected that they must rely on their social network. This finding implies that online community care platforms are not suited for every older adult as such platforms have goals that not every older adult will share; not all older adults need web-based support or require help to find offline support or to request the assistance of others.

Based on our findings, we can conclude that older adults who are still healthy, self-reliant, and internet-skilled seem to be able to find their own means of solving problems and arranging informal care. This is not new because if governments or other organizations that bear responsibility for the welfare of citizens fail to provide people with a sense of security (eg, when state provision is reduced and as a result, older adults are increasingly expected to take care of themselves instead of relying on health care services), people will feel compelled to organize this security themselves. In this scenario, they will look for and find solidarity in small-scale physical groups [19]. It may therefore not be desirable at all to "formally" stimulate or organize local participation or mutual informal care; perhaps, platforms should merely try to provide another means by which citizens themselves can organize local participation and informal care.

Strengths and Limitations of This Study

The strength of this study is that it contributes to knowledge about the (perceived) impact of a web-based community and about its actual use after the implementation. However, we did not intend to demonstrate actual causality between the functionalities of the web-based platform and their intended outcomes (eg, an increase in local participation or feelings of local connectedness) but instead aimed to illustrate how older users perceive the use and impact of an online community care platform. Furthermore, the results should be interpreted with some caution considering the modest sample size and the possible "biased sample:" participants were mainly healthy, highly educated, and internet-skilled. Since the primary target groups of the platform were older adults and vulnerable inhabitants of Grubbenvorst, we did not include the target group of the platform in its entirety, as vulnerable people were not included in the sample. Moreover, as the purpose of this study was to understand the expectations and perceived impact of a relatively small group of users regarding a web-based platform, it might have been of added value to additionally adopt qualitative methods to answer our research questions.

Additionally, it is possible that the "first-come-first-served principle" as adopted in the recruitment process resulted in a biased sample in which users who used Grubbenvorst-Online more than others registered to participate in the study. Perhaps, because of this, the depiction we now have of the use of Grubbenvorst-Online is more positive than actually is the case. Furthermore, the strength of our study would have been improved if we had asked participants the length of time they had been using Grubbenvorst-Online since expectations of new users would possibly be different from the expectations of a long-term user. Finally, in the questionnaire, we did not ask the types of help participants were willing to offer or accept. Thus, we cannot be certain whether older adults do not need support or informal care or whether they have simply already arranged it themselves without the help of a web-based platform or other tool.

Implications for Practice and Research

Previous research demonstrates that various organizations and governments implement online community care platforms with the aim of supporting (vulnerable) citizens [3,5-9]. Based on our findings, we must ask ourselves whether such platforms provide enough "added value" in their current form. Perhaps another direction is desirable, one in which online community care platforms primarily support existing offline initiatives to stimulate local participation, informal caregiving, and feelings of connectedness. An online community care platform would thus never be considered an end in itself, but as a means to achieve certain goals. We believe that online community care platforms can add value when they encompass the following characteristics: (1) the platform's primary objective is to provide access to offline services in a neighborhood, (2) the platform mainly has a facilitating and intermediary role and, for example, connects "help requests" of older adults or (vulnerable) citizens with available local resources (eg, other citizens, volunteers, and professionals from health care organizations), and (3) human representatives of the platform are physically present in the neighborhood during set times, and offline communication (eg, by telephone or face-to-face interactions) is the primary form by which older adults may voice their help requests. In this model, the idea of a "one-stop shop" becomes the guiding principle underlying an online community care platform, that is, all local help requests and local resources from a neighborhood come together in one place.

Additionally, previous research indicates that many (eHealth) apps (such as certain functionalities of the online community care platform) lack a clear theoretical basis and do not provide any evidence concerning their effectiveness and usability [20]. In the further development of online community care platforms, it is advisable to pay more attention to the theoretical "rationale" of the platform and its functionalities. Another important recommendation regarding the future development and implementation of online community care platforms is to involve potential end users. Previous research stresses the importance of including users to constitute web-based communities that will be functional, usable, and accessible [11,21,22]. Finally, regarding the future development of online community care platforms, it is advisable that developers continually rethink the platforms' features, make adjustments, and evaluate whether the functionalities are actually effective.

Further research is required to establish if online community care platforms have an impact on older adults' local participation, informal caregiving, and feelings of connectedness, and if so, the ways in which this impact is felt. Research should focus on the impact of different online community care platforms and on their impact on the group of older adults in all its variety: youngest-old, middle-old, and oldest-old; frail and healthy; highly and poorly educated; and internet-skilled and unskilled. Finally, since platforms often focus on all inhabitants of a neighborhood to create an interdependent support system where old and young people are facilitated to help each other, further work should also focus on researching the impact of these platforms on user groups other than older adults.

Finally, it would be interesting to study the impact of the current coronavirus pandemic on the adoption and use of web-based platforms. Perhaps the pandemic may lead to an increased use of these platforms, as it has led to a striking adaptation of various eHealth services in community care (which used to have a history of strenuous and slow adoption and implementation).

Acknowledgments

The funding for this study (project number 102.515) was provided by the NutsOhra Fund (FNO) and Zuyd University of Applied Sciences. Neither funding body had a role in the collection, analysis, and interpretation of data, nor in writing the manuscript. The authors thank the members of the project group "Meedoen in Grubbenvorst" (Peter Jacobs, John Rietman, Willemien van Gardingen, Ruud Meijers, Mieke Cruijsberg, and Kees Cligge) for their help in recruiting participants and for providing feedback on the questionnaire. Finally, we thank Stephen Potter for checking the manuscript as a native English speaker.

Authors' Contributions

All authors were involved in drafting and revising the manuscript and gave their final approval of the version to be published. They also agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Smits CHM, van den Beld HK, Aartsen MJ, Schroots JJF. Aging in the Netherlands: state of the art and science. Gerontologist 2014 Jun;54(3):335-343. [doi: 10.1093/geront/gnt096] [Medline: 24000267]
2. van Hees SV. Chapter 1. Introduction. In: The making of ageing-in-place: perspectives on a Dutch social policy towards lifecycle-robust neighbourhoods. Maastricht University: Creating Value-Based Health Care; Promovendi PHPC; Health Services Research; 2014:7-22.
3. Boll F, Brune P. Online Support for the Elderly – Why Service and Social Network Platforms should be Integrated. Procedia Computer Science 2016;98:395-400. [doi: 10.1016/j.procs.2016.09.060]
4. Grandcare Systems. Home Monitoring System For Elderly: Technology Enabled Senior Care Solutions For Better Home Health Care Computer For Seniors. URL: https://www.grandcare.com/ [accessed 2020-05-10]
5. De Jong Y. Overzicht Digitale Communicatieplatforms om te communiceren binnen de zorg en de zorg te delen. Zorg voor Beter. 2019 Mar. URL: https://www.zorgvoorbeter.nl/zorgvoorbeter/media/documents/thema/familieparticipatie/digitale-platforms-zorgverleners-familie.pdf [accessed 2019-10-23]
6. Krijgsman J, Eertink H, Leeuw J, Zondervan R. Whitepaper Praktisch model voor ICT-platformen bij welzijn en zorg op afstand. Nictiz. 2012 Jun 19. URL: https://www.nictiz.nl/wp-content/uploads/2012/06/Praktisch-model-voor-ICT-platformen-voor-welzijn-en-zorg-op-afstand-1.0.pdf [accessed 2019-10-23]
7. Makai P, Perry M, Robben SHM, Schers H, Heinen M, Olde Rikkert MGM, et al. Which frail older patients use online health communities and why? A mixed methods process evaluation of use of the Health and Welfare portal. J Med Internet Res 2014 Dec 17;16(12):e278 [FREE Full text] [doi: 10.2196/jmir.3609] [Medline: 25519769]
8. Makai P, Perry M, Robben SHM, Schers HJ, Heinen MM, Olde Rikkert MGM, et al. Evaluation of an eHealth intervention in chronic care for frail older people: why adherence is the first target. J Med Internet Res 2014 Jun 23;16(6):e156 [FREE Full text] [doi: 10.2196/jmir.3057] [Medline: 24966146]
9. Willard S, van Rossum E, Spreeuwenberg M, de Witte L. A typology of online care platforms for community-dwelling older adults in the Netherlands: A scoping review. Gerontechnology 2019 Sep 28;18(3):122-141. [doi: 10.4017/gt.2019.18.3.001.00]

10. Farkas P, Schrenk M, Hlauschek W. Senior Social Platform. An application aimed to reduce the social and digital isolation of seniors. In: REAL CORP. 2010 May 18 Presented at: REAL CORP 2010: CITIES FOR EVERYONE? Liveable, Healthy, Prosperous Promising vision or unrealistic fantasy?; May 18 2010; Vienna p. 1247-1252 URL: https://conference.corp.at/archive/CORP2010_223.pdf
11. Jaeger PT, Bo Xie. Developing Online Community Accessibility Guidelines for Persons With Disabilities and Older Adults. Journal of Disability Policy Studies 2008 Oct 17;20(1):55-63. [doi: 10.1177/1044207308325997]
12. Koene P, Köbler F, Esch S, Leimeister J, Krcmar H. Design and evaluation of a service-oriented collaborative consumption platform for the elderly. : Association for Computing Machinery New York NY United States; 2012 May Presented at: CHI '12: CHI Conference on Human Factors in Computing Systems; May 2012; Austin, Texas, USA p. 2537-2542. [doi: 10.1145/2212776.2223832]
13. Marcelino I, Laza R, Pereira A. SSN: Senior Social Network for Improving Quality of Life. International Journal of Distributed Sensor Networks 2016 Jul 22;12(7):2150734. [doi: 10.1177/155014772150734]
14. Peek STM, Wouters EJM, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJM. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
15. Willard S, Cremers G, Man YP, van Rossum E, Spreeuwenberg M, de Witte L. Development and testing of an online community care platform for frail older adults in the Netherlands: a user-centred design. BMC Geriatr 2018 Apr 07;18(1):87 [FREE Full text] [doi: 10.1186/s12877-018-0774-7] [Medline: 29625562]
16. Kerncijfers wijken en buurten 2017. CBS Statline. 2017. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83765ned/table?fromstatweb [accessed 2019-06-10]
17. Cools-Tummers G, Van Zundert P. Seniorenmonitor 2015: De behoeften van Heerlense senioren in kaart gebracht. Municipality of Heerlen. Heerlen; 2015. URL: http://www.bgwheerlen.nl/wp-content/uploads/2013/07/Seniorenmonitor_2015.pdf [accessed 2019-08-22]
18. Frieling M. Een goede buur: 'Joint production' als motor voor actief burgerschap in de buurt. In: Rijksuniversiteit Groningen. Groningen: Rijksuniversiteit Groningen; 2018.
19. Aslander M, Witteveen E. Een Beetje Nooit Af. Amsterdam/Antwerpen: Business Contact; 2016.
20. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone Apps to Support Self-Management of Hypertension: Review and Content Analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
21. Nef T, Ganea RL, Müri RM, Mosimann UP. Social networking sites and older users - a systematic review. Int Psychogeriatr 2013 Jul;25(7):1041-1053. [doi: 10.1017/S1041610213000355] [Medline: 23552297]
22. Peek STM, Wouters EJM, Luijkx KG, Vrijhoef HJM. What it Takes to Successfully Implement Technology for Aging in Place: Focus Groups With Stakeholders. J Med Internet Res 2016 May 03;18(5):e98 [FREE Full text] [doi: 10.2196/jmir.5253] [Medline: 27143097]

Edited by G Eysenbach; submitted 15.05.20; peer-reviewed by J Kim, R Cruz Martínez, D Carvalho; comments to author 08.07.20; revised version received 04.09.20; accepted 26.10.20; published 01.12.20.

Please cite as:
Willard S, van Rossum E, Spreeuwenberg M, de Witte L
Perceived Impact of an Online Community Care Platform for Dutch Older Adults on Local Participation, Informal Caregiving, and Feelings of Connectedness: Pretest-Posttest Observational Study
J Med Internet Res 2020;22(12):e20304
URL: https://www.jmir.org/2020/12/e20304
doi:10.2196/20304
PMID:33258779

Original Paper

Association Rule Mining and Prognostic Stratification of 2-Year Longevity in Octogenarians Undergoing Endovascular Therapy for Lower Extremity Arterial Disease: Observational Cohort Study

Jing-Yi Jhang[1*], MD; I-Shiang Tzeng[2*], PhD; Hsin-Hua Chou[1,3], MD; Shih-Jung Jang[1,3], MD; Chien-An Hsieh[1], MD; Yu-Lin Ko[1,3], PhD; Hsuan-Li Huang[1,4], MD

[1]Division of Cardiology, Department of Internal Medicine, Taipei Tzu-Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei, Taiwan
[2]Department of Research, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei, Taiwan
[3]School of Medicine, Tzu Chi University, Hualien, Taiwan
[4]School of Post-Baccalaureate Chinese Medicine, Tzu Chi University, Hualien, Taiwan
[*]these authors contributed equally

Corresponding Author:
Hsuan-Li Huang, MD
Division of Cardiology
Department of Internal Medicine
Taipei Tzu-Chi Hospital, Buddhist Tzu Chi Medical Foundation
289 Jiang Kuo Road
Xindian District
New Taipei, 231
Taiwan
Phone: 886 2 66289779 ext 8655
Fax: 886 2 66289009
Email: huang304@gmail.com

Abstract

Background: Two-year longevity is a crucial consideration in revascularization strategies for patients with symptomatic lower extremity arterial disease (LEAD). However, factors associated with 2-year longevity and risk stratification in octogenarians or nonagenarians have been underreported.

Objective: This paper aims to investigate the associated variables and stratify the 2-year prognosis in older patients with LEAD.

Methods: We performed logistic regression and association rule mining based on the Apriori algorithm to discover independent variables and validate their associations with 2-year longevity. Malnutrition, inflammation, and stroke factors were identified. C statistics and Kaplan-Meier analysis were used to assess the impact of different numbers of malnutrition, inflammation, and stroke factors on 2-year longevity.

Results: We recruited a total of 232 octogenarians or nonagenarians (mean age 85 years, SD 4.2 years) treated with endovascular therapy. During the study period, 81 patients died, and 27 of those (33%) died from a cardiac origin within 2 years. Association rules analysis showed the interrelationships between 2-year longevity and the neutrophil-lymphocyte ratio (NLR) and nutritional status as determined by the Controlling Nutritional Status (CONUT) score or Geriatric Nutritional Risk Index (GNRI). The cut-off values of NLR, GNRI, and CONUT were ≥3.89, ≤90.3, and >3, respectively. The C statistics for the predictive power for 2-year longevity were similar between the CONUT score and the GNRI-based models (0.773 vs 0.760; *P*=.57). The Kaplan-Meier analysis showed that 2-year longevity was worse as the number of malnutrition, inflammation, and stroke factors increased from 0 to 3 in both the GNRI-based model (92% vs 68% vs 46% vs 12%, respectively; *P*<.001) and the CONUT score model (87% vs 75% vs 49% vs 10%, respectively; *P*<.001). The hazard ratio between those with 3 factors and those without was 18.2 (95% CI 7.0-47.2; *P*<.001) in the GNRI and 13.6 (95% CI 5.9-31.5; *P*<.001) in the CONUT score model.

Conclusions: This study demonstrated the association and crucial role of malnutrition, inflammation, and stroke factors in assessing 2-year longevity in older patients with LEAD. Using this simple risk score might assist clinicians in selecting the appropriate treatment.

(J Med Internet Res 2020;22(12):e17487) doi:10.2196/17487

KEYWORDS

endovascular therapy; lower extremity arterial disease; octogenarians; longevity; association rules; older people; prognosis; risk; medical informatics; clinical informatics

Introduction

With the aging of the population and improvements in the quality of medical care, physicians encounter an ever-increasing number of older patients with advanced forms of lower extremity arterial disease (LEAD) [1,2]. The results of the Bypass Versus Angioplasty in Severe Ischaemia of the Leg (BASIL) study [3] and the American Heart Association and American College of Cardiology guidelines [4] suggested bypass surgery as an appropriate first-line revascularization procedure for chronic limb-threatening ischemia (CLTI) in patients with a life expectancy of more than 2 years. However, advanced age is associated with increased perioperative and postoperative mortality after vascular operations because of the presence of multiple comorbidities [5-7]. Some studies have reported that endovascular therapy (EVT) in older patients is safe and effective for lifestyle-limiting intermittent claudication and could be an effective alternative treatment for CLTI [1,6,8,9].

Although 2-year longevity is a crucial consideration in revascularization strategies in patients with symptomatic LEAD, this measure should be used with caution in older people with an inherently shorter life expectancy, and the factors associated with 2-year longevity and the prognostic stratification in these patients are unclear.

Apriori algorithm–based association rules analysis (ARA) is widely used to analyze interesting statistical correlations hidden in sets of multidimensional data and might facilitate the process of disease management [10-12].

The aim of this study was to investigate the factors predicting 2-year life expectancy in octogenarians or nonagenarians with LEAD in order to help clinicians care for these patients.

Methods

Study Design

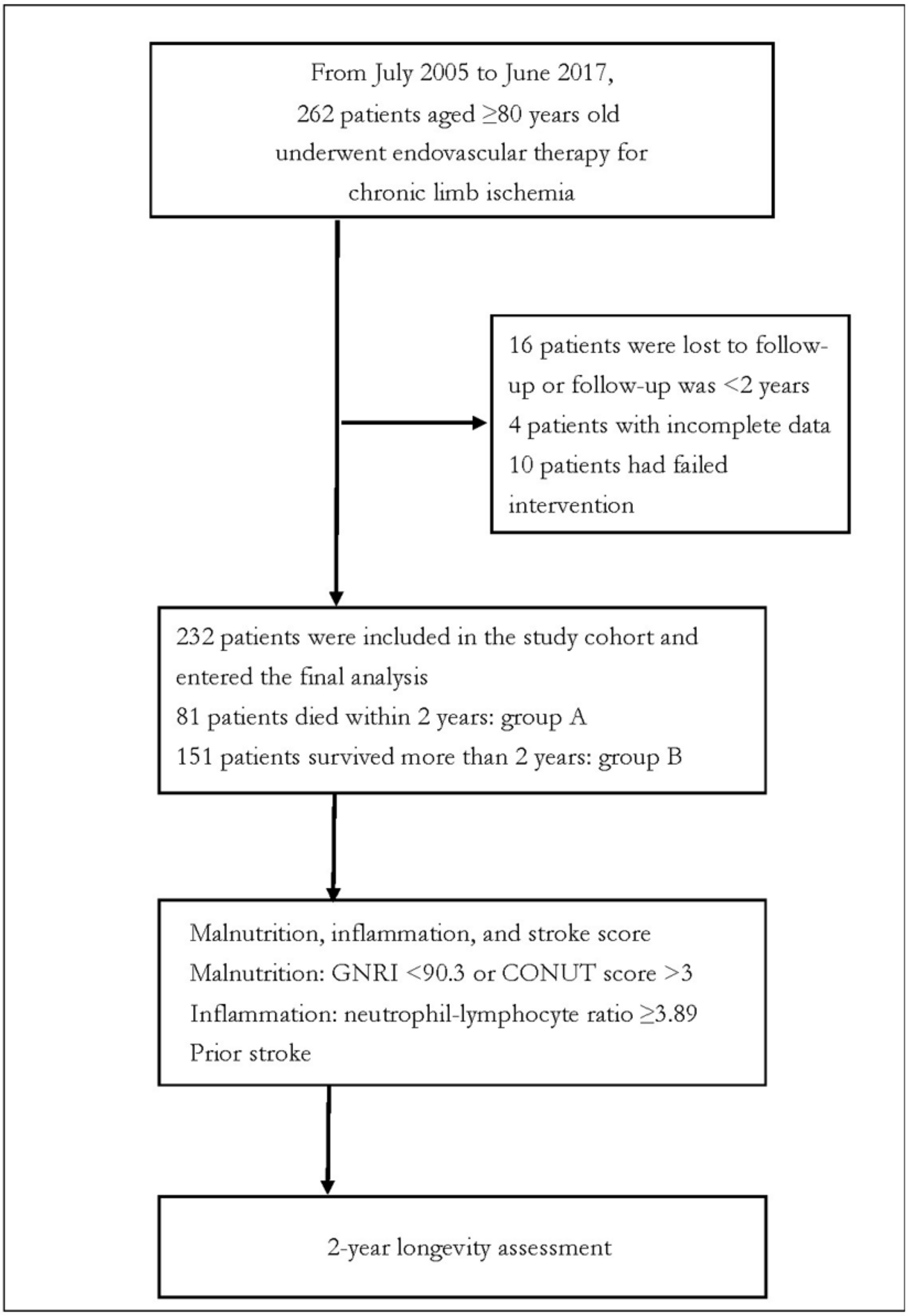
This was a single-center observational cohort study and was approved by the institutional review board (approval No. 06-X18-098) of Taipei Tzu-Chi Hospital, New Taipei City, Taiwan. We obtained informed consent from all participants.

Study Population

Data were extracted from the Tzuchi Registry of Endovascular Intervention for Peripheral Artery Disease, which is a single-center observational registry of patients who have undergone EVT for LEAD starting from July 2005. This retrospective investigation included patients older than 80 years with symptomatic LEAD who were treated between July 2005 and June 2017.

Patients were considered eligible for enrollment if they were candidates for EVT for atherosclerotic LEAD and had given their consent to participate in the study. Excluded from the study were patients with nonatherosclerotic LEAD, acute limb ischemia, overwhelming foot infection, and a follow-up duration of less than 2 years. Figure 1 depicts patients' enrollment in this study. All study procedures were in line with the principles outlined in the Helsinki Declaration.

Figure 1. Flowchart of the inclusion of the study participants. CONUT: Controlling Nutritional Status; GNRI: Geriatric Nutritional Risk Index.



The detailed definitions of demographic variables and the pre-EVT assessments, including clinical examination, hemodynamic evaluation, and anatomical assessment of the target limb, have been described previously [8,13].

Interventions

Endovascular therapy was performed by experienced interventional cardiologists according to the Trans-Atlantic Inter-Society Consensus II guideline recommendations [14]. The detailed procedures of EVT have also been described previously [8,13]. Most of the study participants underwent plain balloon angioplasty alone or bare-metal nitinol stent implantation. With the advancement in technology and the invention of new devices, atherectomy and drug-eluting devices have been introduced to optimize the angiographic results and improve vessel patency. Antiplatelet therapy, anticoagulant regimens, and other medications after EVT depended on the physician’s discretion and the patient’s condition.

Biochemical Markers

Blood samples were obtained at admission or before EVT for the measurement of serum albumin, fasting blood sugar, glycohemoglobin, C-reactive protein (CRP), complete blood count (CBC), and differential counts for the determination of immune-inflammatory variables; this included the neutrophil-lymphocyte ratio (NLR), the platelet-lymphocyte ratio (PLR), and the systemic immune-inflammation index (SII), defined as (platelet count × neutrophil count)/lymphocyte count. The total cholesterol, triglyceride, and high- and low-density lipoprotein cholesterol levels were measured at admission, and the most recent values obtained within 2 months before the procedure were also examined.

Nutritional Status

The Geriatric Nutritional Risk Index (GNRI) and the Controlling Nutritional Status (CONUT) score are widely used in various clinical settings in the field of malnutrition [15,16]. They have been reported as prognostic predictors in patients with LEAD [17]. The CONUT score includes serum albumin, total cholesterol levels, and total lymphocyte count. CONUT scores of 0 to 1 indicate a normal nutritional status, scores of 2 to 4 indicate mild risk, scores of 5 to 8 indicate moderate risk, and scores of 9 to 12 indicate severe risk of malnutrition. GNRI was also calculated using the following equation: GNRI = (14.89 serum albumin [g/dL]) + (41.7 × [body weight/ideal body weight]) [18].

Association Rule Analysis and Grouped Matrix Method

ARA was initially designed to detect and extract useful information from large-scale business databases [10,11]. Recently, this method has been widely applied in clinical medicine to facilitate disease diagnosis and prevention and improve the survival rate [19,20]. The fundamental concept of ARA usually takes the form of $A \rightarrow B$, which means A implies B for any set of variables A and B. Briefly, the implication of ARA is co-occurrence and not causality. For clinical applications, the frequent association between combinations of variables is used to determine the expected prevalence (or occurrence) of disease over random chance. Three core values are used to determine interesting or useful rules: support, confidence, and lift.

Support is defined as P(A,B) (ie, the joint probability of A and B), that is, the prevalence of both A and B co-occurring. Confidence is the conditional probability that B occurs, given that A is already present. The lift (presented as the observed to expected ratio) is expressed as P(A,B) to P(A) P(B), which is a measure of the level of dependence between A and B. A lift of 1 indicates that A and B are independent without any association with each other. In ARA, the sets of A and B are restricted to a single variable because many possible combinations of associations exist between variables (122 in this study). The grouped matrix method creates nested groups of rules via clustering. The nested groups form a hierarchy, which can interactively explore multiple variables of rules that precede the following variable.

Definitions and Outcome End Points

We defined procedural success as the successful restoration of blood flow, with evidence of at least one preexisting or reestablished crural vessel to the foot. The lesion score is the sum of the scores of the diseased lower-extremity blood vessels. A score of 1 point each was given for the iliac, femoropopliteal, anterior tibial, posterior tibial, and peroneal arteries [21].

Nonambulatory patients were defined as patients who used wheelchairs or patients with a bedridden status. Cardiovascular death included sudden cardiac arrest or death caused by myocardial infarction, stroke, lethal arrhythmia, decompensated heart failure, valvular heart disease, and aortic or other vascular diseases.

The outcome end point was 2-year survival with risk stratification according to the determining factors of 2-year longevity.

Follow-up

After EVT, we performed scheduled follow-ups at 1 month and every 3 months with clinical assessment and duplex ultrasound evaluations. Repeat EVT was performed if symptoms recurred. Major events (death, amputation, and all vascular events) were documented at the follow-up visits. If data of office follow-up visits were not available, alternate data sources included telephone interviews, medical records, and the referring physician. The follow-up closing date was December 31, 2018.

Statistical Analysis

Statistical analysis was performed using SPSS Statistics (version 22.0; IBM Corp) and MedCalc statistical software (version 18.11.3; MedCalc Software). Descriptive statistics were presented as frequency (percent) for categorical variables and mean (standard deviation) for continuous variables. Discrete and categorical data were analyzed using the Pearson chi-square test. The parametric continuous variables were statistically analyzed and compared between groups using an independent *t* test, whereas the Mann-Whitney U test was used to analyze nonparametric continuous and ordinal data. CBC, NLR, PLR, SII, and CRP levels were presented as medians and interquartile ranges and were logarithmically transformed before statistical analysis. A receiver operating characteristic (ROC) curve was used to determine the cut-off values of the NLR, GNRI, and CONUT score for 2-year longevity prediction. The area under the curve (AUC) of these variables was compared using the DeLong method. Cox proportional hazard model analysis was performed to identify the independent predictors associated with 2-year longevity. The predictive performance levels of the combination model for 2-year longevity in C statistics were also compared using the DeLong method. A Kaplan-Meier analysis was conducted to compare 2-year longevity using the number of malnutrition, inflammation, and stroke factors. *P* values of $<.05$ were considered statistically significant.

We validated the associations of malnutrition and inflammation with 2-year longevity using ARA [10,11,22] and by visualizing associations using R software (version 3.4.3; The R Foundation for Statistical Computing). The procedure can be conveniently fitted using the R package “arules.” The visualizing association rules can be directly applied using the R package “arulesViz.” A total of 15 potential variables were investigated for association rules, as determined by the minimum requirements of a support degree of ≥20% and confidence of ≥80% in this study.

Results

We treated 262 older patients with EVT from July 2005 to June 2017. Of these, 30 were not eligible for analysis, 16 were lost to follow-up or had a follow-up duration of less than 2 years, 4 had incomplete data, and 10 had failed EVT. The remaining 232 patients were divided into 2 groups based on whether they had a survival duration of ≥2 years. Group A included 81 patients who died (27/81, 33% died from cardiac origin) within 2 years after the index EVT, while group B was made up of 151 patients who survived for more than 2 years.

Table 1 summarizes the baseline demographics of the 2 groups. Patients in group A had a significantly higher incidence of congestive heart failure (CHF), cerebrovascular accident (CVA), dialysis, CLTI, and nonambulatory status than those in group B. BMI, serum albumin level, cholesterol level, and GNRI were significantly higher in group B patients. Higher levels of inflammatory markers (white blood cell [WBC] count, NLR, PLR, SII, and CRP level) and CONUT scores were more frequently observed in group A than in group B.

Table 2 shows the lesion characteristics of the treated limbs in the 2 groups. Group A patients had a lower ankle-brachial pressure index and a higher incidence of CLTI than group B patients. There were no differences between the 2 groups in lesion distribution, complexity, isolated or multilevel EVTs, and stenting rate.

Table 1. Patient demographics.

Factors	All patients	Group A (2-year death)	Group B (2-year survival)	*P* value
Patients, N	232	81	151	N/A[a]
Age (years), mean (SD)	85.4 (4.2)	85.8 (4.0)	85.3 (4.3)	.41
Sex (male), n (%)	109 (47)	36 (44)	73 (48)	.57
Diabetes mellitus, n (%)	130 (56)	51 (62)	79 (52)	.12
Hypertension, n (%)	201 (87)	72 (89)	129 (85)	.46
Coronary artery disease, n (%)	93 (40)	32 (40)	61 (40)	.90
Congestive heart failure, n (%)	39 (17)	19 (24)	20 (13)	.047
Cerebrovascular accident, n (%)	52 (22)	29 (36)	22 (15)	<.001
Dialysis dependence, n (%)	57 (25)	27 (33)	30 (20)	.02
Smoking history, n (%)	58 (25)	19 (24)	39 (26)	.69
Atrial fibrillation, n (%)	49 (21)	20 (25)	29 (19)	.33
Hyperlipidemia, n (%)	95 (41)	31 (39)	64 (42)	.59
Ambulatory status, n (%)	83 (36)	16 (20)	67 (44)	<.001
CLTI[b], n (%)	192 (83)	78 (96)	114 (76)	<.001
Body mass index (kg/m^2), mean (SD)	23.1 (3.5)	22.3 (3.5)	23.6 (3.4)	.006
Cholesterol (mg/dL), mean (SD)	159 (39)	150 (39)	163 (39)	.01
Triglyceride (mg/dL), mean (SD)	117 (70)	111 (63)	119 (73)	.46
Glycohemoglobin (%), mean (SD)	6.51 (1.50)	6.70 (1.68)	6.42 (1.41)	.21
Hematocrit (%), median (IQR)	33.7 (29.8-37.2)	33.5 (29.9-36.5)	33.9 (29.8-38.1)	.10
White blood cell count (10^9/L), median (IQR)	7.230 (5.673-9.080)	7.970 (6.720-9.965)	6.850 (5.120-8.200)	.001
Platelet count (10^3/μL), median (IQR)	208 (164-255)	217 (169-278)	205 (156-245)	.07
Neutrophil-lymphocyte ratio, median (IQR)	3.62 (2.34-5.62)	5.03 (3.31-7.02)	3.03 (2.04-4.64)	<.001
Platelet-lymphocyte ratio, median (IQR)	145 (114-219)	173 (120-243)	135 (107-195)	.002
Systemic immune-inflammation index[c], median (IQR)	716 (461-1272)	1083 (568-1835)	598 (383-972)	<.001
C-reactive protein (mg/dL), median (IQR)	1.25 (0.31-4.40)	3.08 (0.77-8.20)	0.90 (0.19-2.92)	<.001
Albumin (g/dL), mean (SD)	3.10 (0.65)	2.78 (0.63)	3.26 (0.59)	<.001
CONUT[d] score, mean (SD)	4.74 (2.97)	6.26 (2.81)	3.93 (2.72)	<.001
Geriatric Nutritional Risk Index, mean (SD)	89.3 (12.5)	83.1 (12.3)	92.6 (11.3)	<.001
Follow-up time (days), median (IQR)	971 (389-1575)	229 (79-438)	1360 (1008-1942)	<.001

[a] N/A: not applicable.

[b] CLTI: chronic limb-threatening ischemia.

[c] Systemic immune-inflammation index was defined as (platelet count × neutrophil count)/lymphocyte count.

[d] CONUT: Controlling Nutritional Status.

Table 2. Lesion and interventional procedure characteristics.

Factors	All limbs	Group A (2-year death)	Group B (2-year survival)	*P* value
Limbs, n	232	81	151	N/A[a]
Claudication, n (%)	40 (17)	3 (4)	37 (25)	.001
Resting pain, n (%)	51 (22)	19 (23)	32 (21)	N/A
Unhealing ulcer, n (%)	109 (47)	44 (54)	65 (43)	N/A
Gangrene, n (%)	32 (14)	15 (19)	17 (11)	N/A
ABI[b] of affected limbs, mean (SD)	0.54 (0.33)	0.51 (0.37)	0.55 (0.31)	.40
ABI of affected limbs excluding ≥1.4, mean (SD)	0.45 (0.19)	0.41 (0.18)	0.47 (0.19)	.02
Iliac lesions, n (%)	30 (13)	9 (11)	21 (14)	.98
Femoropopliteal lesions , n (%)	167 (72)	61 (75)	106 (70)	.35
Below-the-knee lesions, n (%)	170 (73)	61 (75)	109 (73)	.63
Poor runoff, n (%)	192 (84)	67 (85)	125 (83)	.77
Stenting, n (%)	112 (48)	37 (46)	75 (50)	.53
Lesion score, mean (SD)	3.18 (0.98)	3.30 (0.97)	3.11 (0.98)	.16

[a] N/A: not applicable.

[b] ABI: ankle-brachial index.

Compared with WBCs, SII, and PLR, NLR had the largest AUC (0.648, 0.692, 0.621, and 0.703, respectively) (Multimedia Appendix 1) and was the only significant inflammatory variable for 2-year longevity estimation after multivariate analysis. The cut-off value of NLR for 2-year longevity was 3.89. The AUCs for 2-year longevity estimation by the CONUT score or GNRI were statistically significant (*P*<.001); however, there was no difference between the 2 measures (0.725 vs 0.722; *P*=.90). The cut-off values for the GNRI and the CONUT score for 2-year longevity were 90.3 and >3, respectively. Detailed patient demographics and lesion characteristics are provided in Multimedia Appendices 2 and 3.

Table 3 shows the results of 2 separate multivariate analyses examining the predictors of 2-year longevity (model 1 was adjusted for ambulatory status, CHF, CVA, CLTI, dialysis, NLR, and total cholesterol; model 2 was adjusted for the covariates, except that BMI was used to replace total cholesterol). CVA, NLR, GNRI, and the CONUT score were independent factors of 2-year longevity.

Table 3. Results of multivariate logistic regression analysis.

Variables	Model 1 (GNRI[a] + NLR[b])		Model 2 (CONUT[c] + NLR)	
	HR[d] (95% CI)	*P* value	HR (95% CI)	*P* value
Ambulatory status	0.718 (0.291-1.773)	.47	0.776 (0.320-1.884)	.58
Congestive heart failure	2.167 (0.950-4.945)	.07	1.932 (0.827-4.513)	.13
Cerebrovascular accident	2.763 (1.231-6.202)	.01	2.577 (1.169-5.678)	.02
C-reactive protein	1.135 (0.890-1.149)	.31	1.193 (0.933-1.527)	.16
CLTI[e]	1.961 (0.425-9.043)	.39	2.439 (0.535-11.110)	.25
Dialysis	1.774 (0.816-3.855)	.15	1.711 (0.807-3.626)	.16
Cholesterol	1.007 (0.998-1.016)	.16	N/A[f]	N/A
Body mass index	N/A	N/A	1.087 (0.979-1.206)	.12
CONUT score >3	N/A	N/A	2.718 (1.217-6.073)	.02
NLR >3.89	2.679 (1.312-5.470)	.007	2.532 (1.236-5.187)	.01
GNRI <90.3	3.071 (1.447-6.516)	.003	N/A	N/A

[a] GNRI: Geriatric Nutritional Risk Index.
[b] NLR: neutrophil-lymphocyte ratio.
[c] CONUT: Controlling Nutritional Status.
[d] HR: hazard ratio.
[e] CLTI: chronic limb-threatening ischemia.
[f] N/A: not applicable.

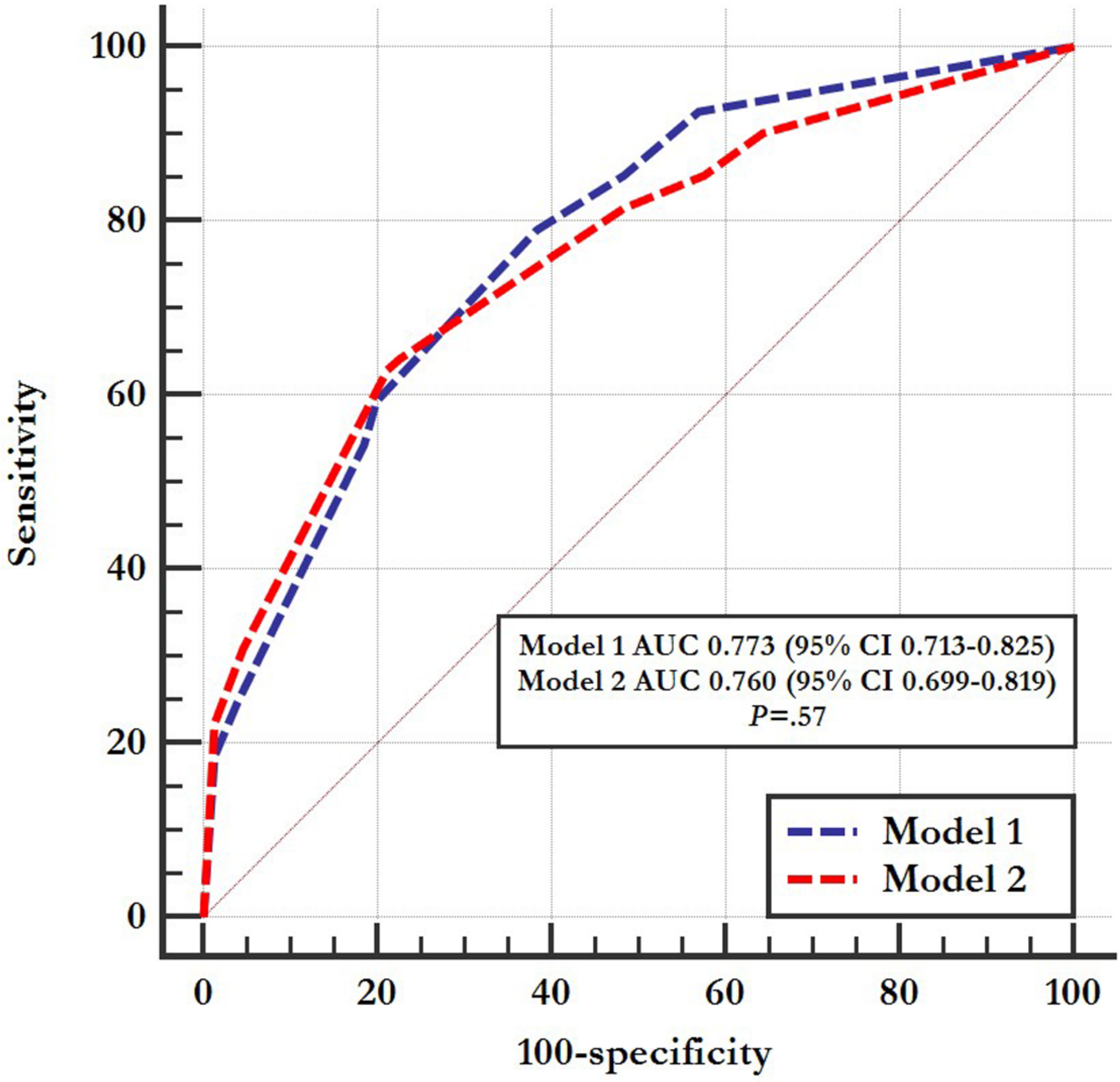
Table 4 shows the C statistics indicating the predictive performance level for 2-year longevity. The C statistic was 0.685 for the GNRI and 0.666 for the CONUT score initially. When CVA and NLR >3.89 were added to the regression model, it resulted in significant stepwise improvements of the C statistic from 0.685 to 0.773 in model 1 (*P*=.006) and 0.666 to 0.760 in model 2 (*P*=.004). However, adding CHF did not further improve the predictive value for 2-year longevity. The predictive performance levels for 2-year longevity was similar between the 2 models (0.773 vs 0.760; *P*=.57) (Figure 2).

Table 4. C statistics for the prediction of 2-year longevity.

Variable	Model 1		Model 2	
	C statistic (95% CI)	*P* value[a]	C statistic (95% CI)	*P* value[a]
GNRI[b] <90.3	0.685 (0.621-0.744)	N/A[c]	N/A	N/A
CONUT[d] score >3	N/A	N/A	0.666 (0.601-0.726)	N/A
Addition of CVA[a,e]	0.736 (0.675-0.792)	<.001	0.723 (0.660-0.779)	<.001
Further addition of NLR[a,f] >3.89	0.773 (0.713-0.825)	.006	0.760 (0.699-0.813)	.004
Further addition of CHF[a,g]	0.774 (0.714-0.826)	.91	0.758 (0.698-0.812)	.86

[a] Compared with the previous model.
[b] GNRI: Geriatric Nutritional Risk Index.
[c] N/A: not applicable.
[d] CONUT: Controlling Nutritional Status.
[e] CVA: cerebrovascular accident.
[f] NLR: neutrophil-lymphocyte ratio.
[g] CHF: congestive heart failure.

Figure 2. Receiver operating characteristic curves of the malnutrition, inflammation, and stroke model for 2-year survival prediction. AUC: area under the curve.



We investigated 122 association rules based on 2-year longevity data. The top 10 association rules (Multimedia Appendix 4) are visually presented on a scatter plot (Figure 3). All rules with a high lift can be easily identified. Rules with a high lift typically have low support. The most interesting rules (support-confidence optimal rules) reside on the support-confidence border, which can be clearly seen in this plot. The association rules between 15 potential variables are ordered by support. Figure 4 shows the presented features on a grouped matrix of 10 associations. The following sets of items were interactively selected to reveal the preceding variables of rules and the following variable based on a grouped matrix for 10 rules: {GNRI, 2-year longevity} → {CONUT} and {NLR, 2-year longevity} → {lnWBC}. These data sets illustrated the close associations of malnutrition (GNRI, CONUT) and inflammation (NLR) factors with 2-year longevity in older patients.

Figure 3. Top 10 association rules. The figure shows the plot with lift on the y-axis.

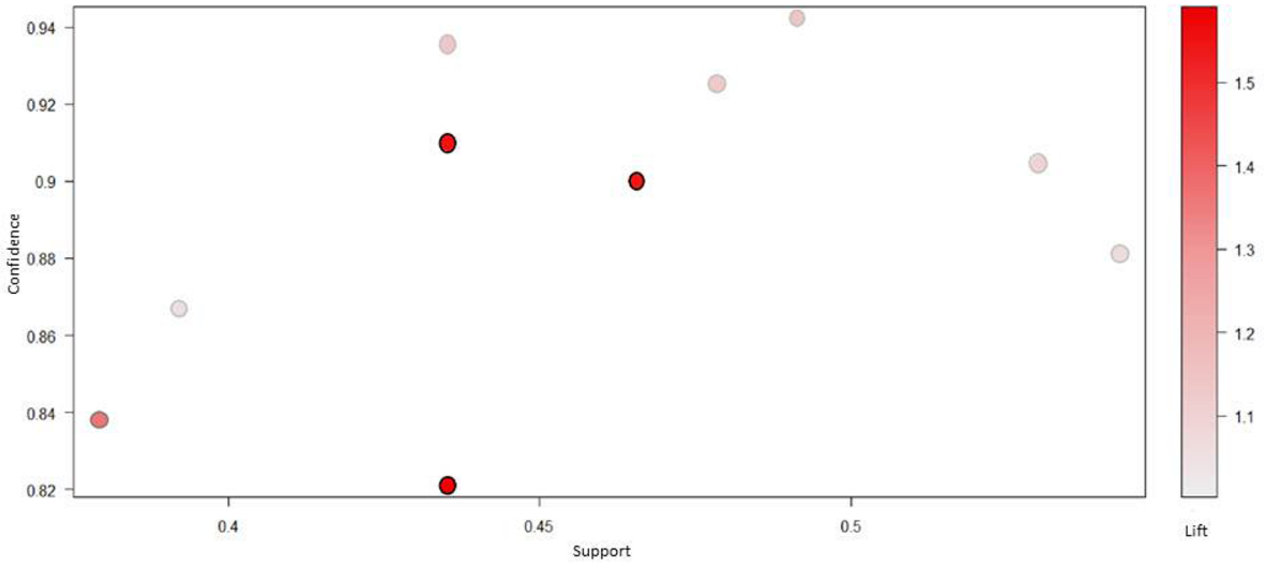
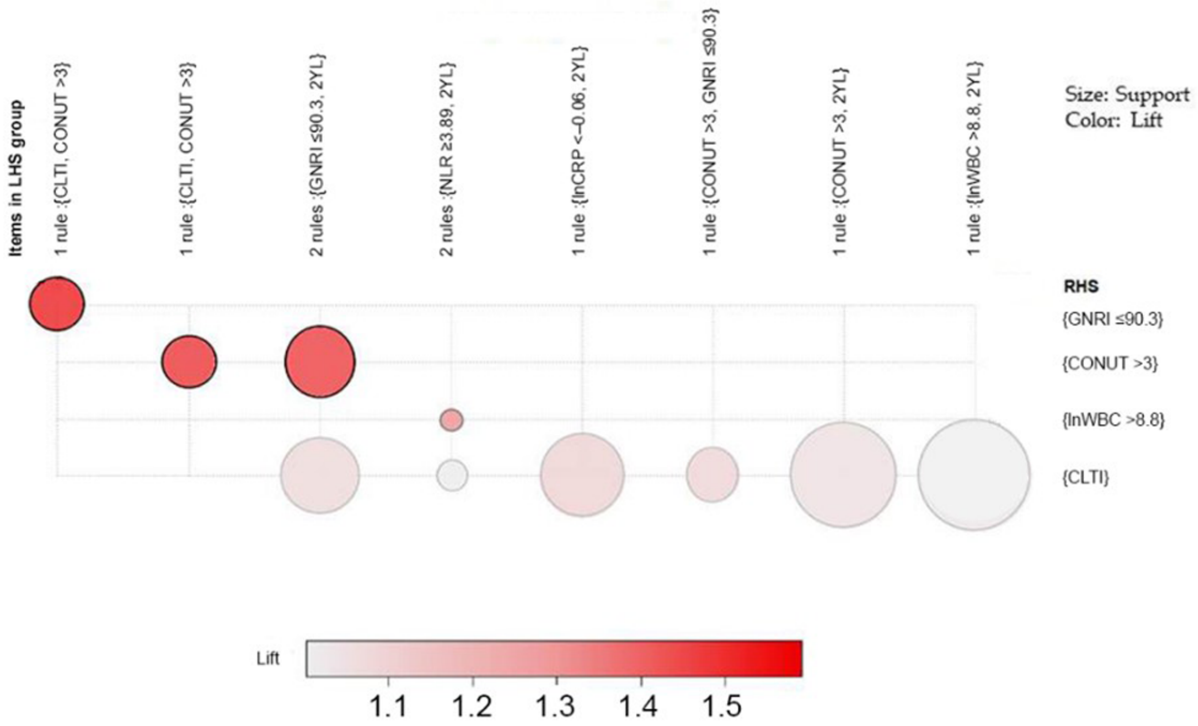
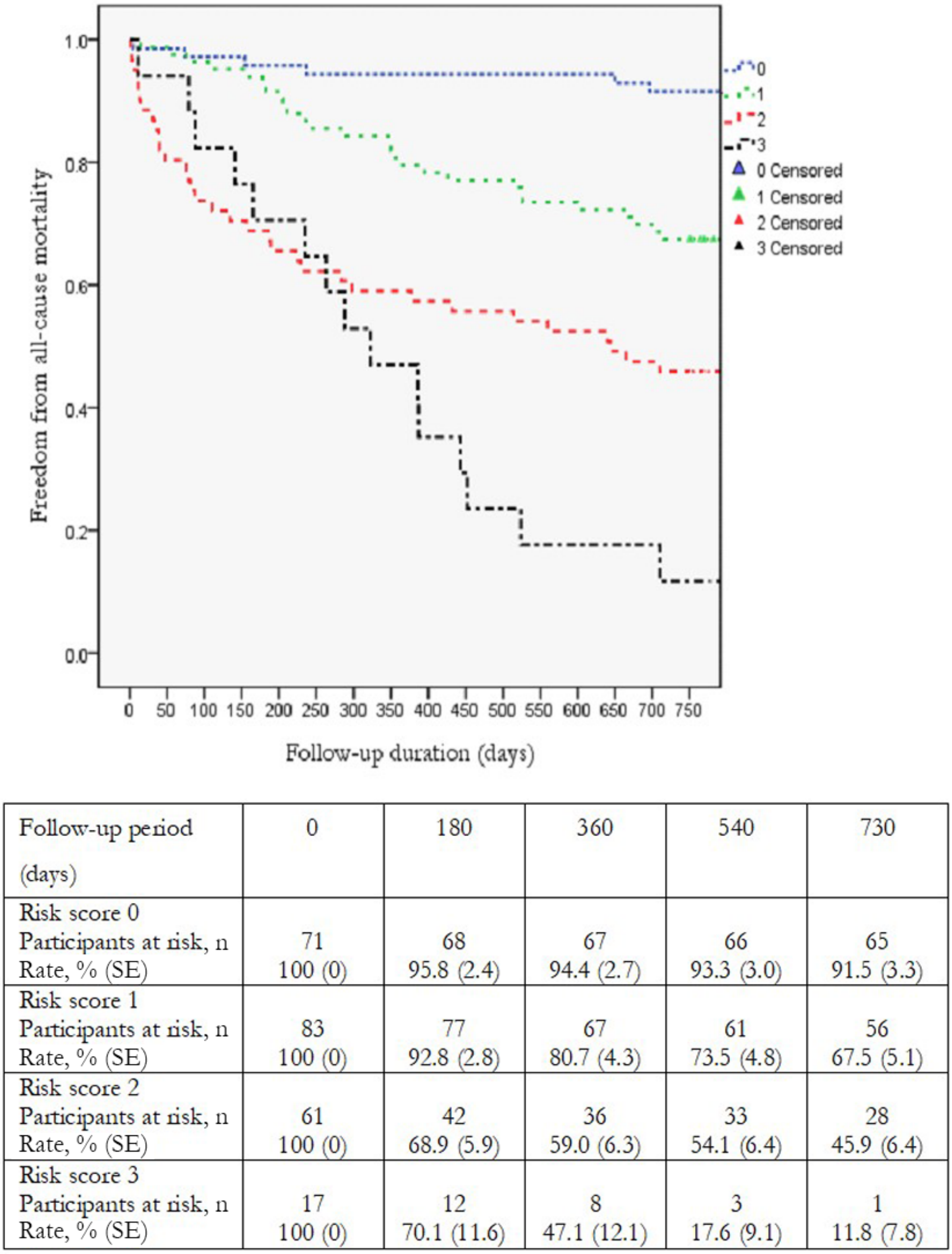


Figure 4. Grouped matrix for 10 rules. The graph-based visualization uses color or size to represent the set of items and rules. This graph offers a representation of rules and enables a small set of rules to avoid a cluttered presentation. The left-hand side shows the antecedents and the right-hand side shows the consequents. 2YL: 2-year longevity; CLTI: chronic limb-threatening ischemia; CONUT: Controlling Nutritional Status; CRP: C-reactive protein; GNRI: Geriatric Nutritional Risk Index; LHS: left-hand side; NLR: neutrophil-lymphocyte ratio; RHS: right-hand side; WBC: white blood cell.



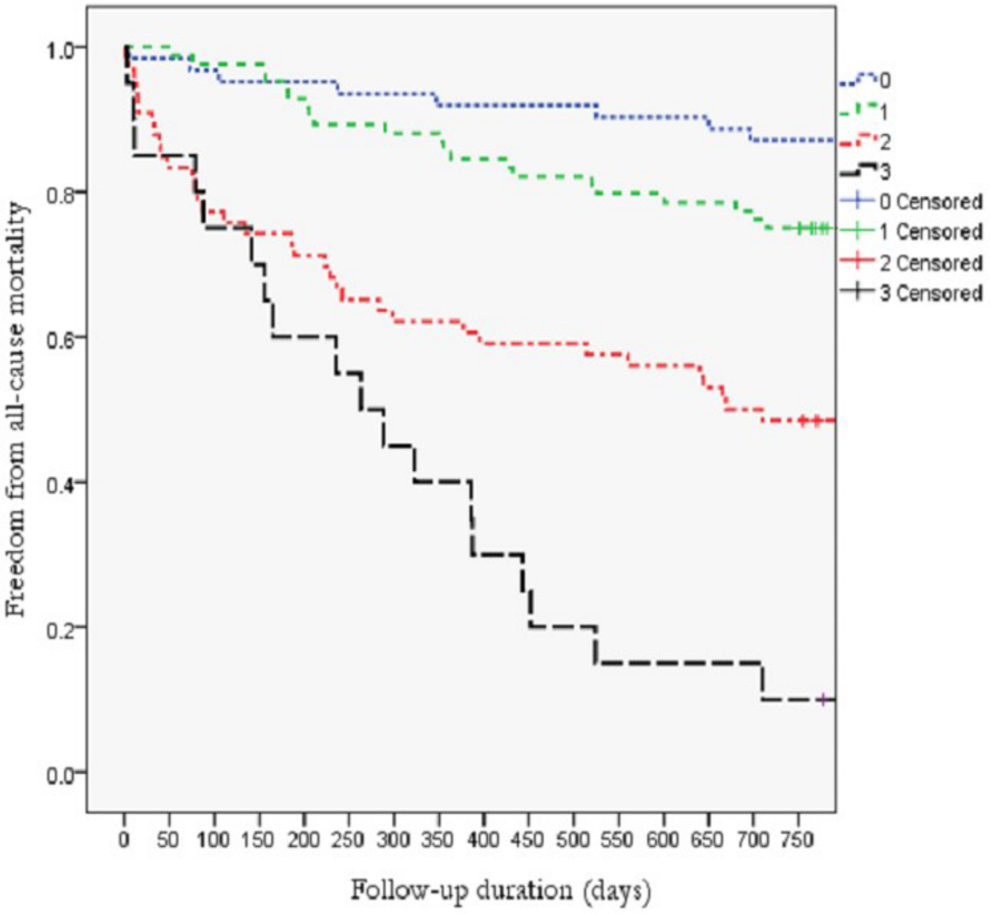
Kaplan-Meier analysis demonstrated a significant decrease in 2-year survival as the number of malnutrition, inflammation, and stroke factors increased from 0 to 3 in both the GNRI-based model (92% vs 68% vs 46% vs 12%, respectively; *P*<.001) and the CONUT score model (87% vs 75% vs 49% vs 10%, respectively; *P*<0.001) (Figures 5 and 6), which was in line with the hazard ratio (HR) between patients with 3 factors and those without (HR 18.2, 95% CI 7.0-47.2; *P*<.001 in the GNRI model and HR 13.6, 95% CI 5.9-31.5; *P*<.001 in the CONUT model).

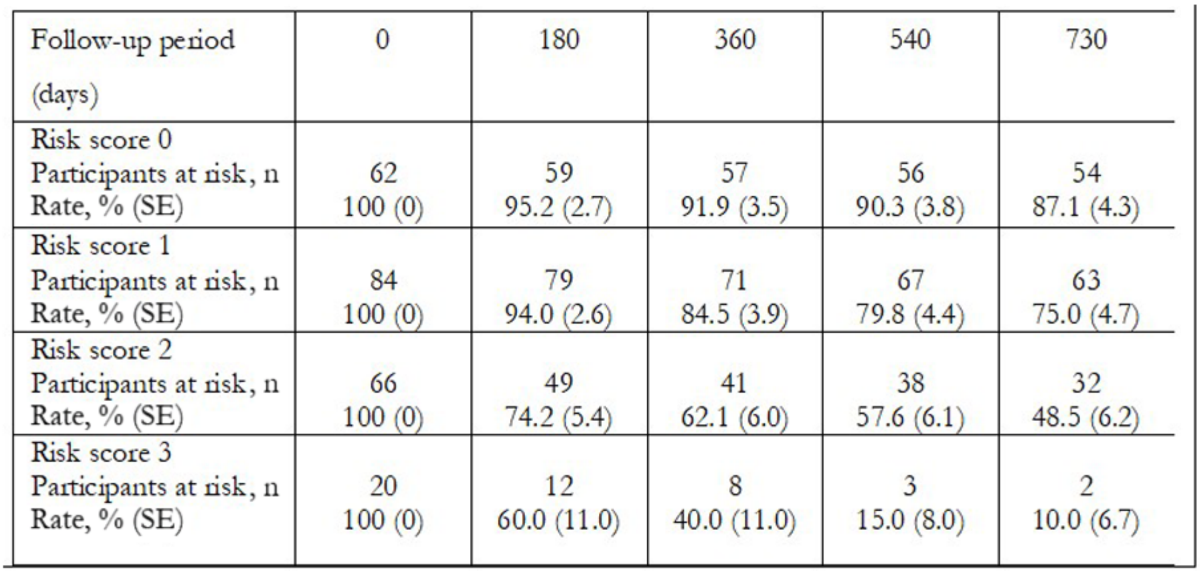
Figure 5. Kaplan-Meier analysis of 2-year survival in model 1.



Follow-up period (days)	0	180	360	540	730
Risk score 0					
Participants at risk, n	71	68	67	66	65
Rate, % (SE)	100 (0)	95.8 (2.4)	94.4 (2.7)	93.3 (3.0)	91.5 (3.3)
Risk score 1					
Participants at risk, n	83	77	67	61	56
Rate, % (SE)	100 (0)	92.8 (2.8)	80.7 (4.3)	73.5 (4.8)	67.5 (5.1)
Risk score 2					
Participants at risk, n	61	42	36	33	28
Rate, % (SE)	100 (0)	68.9 (5.9)	59.0 (6.3)	54.1 (6.4)	45.9 (6.4)
Risk score 3					
Participants at risk, n	17	12	8	3	1
Rate, % (SE)	100 (0)	70.1 (11.6)	47.1 (12.1)	17.6 (9.1)	11.8 (7.8)

Figure 6. Kaplan-Meier analysis of 2-year survival in model 2.





Follow-up period (days)	0	180	360	540	730
Risk score 0					
Participants at risk, n	62	59	57	56	54
Rate, % (SE)	100 (0)	95.2 (2.7)	91.9 (3.5)	90.3 (3.8)	87.1 (4.3)
Risk score 1					
Participants at risk, n	84	79	71	67	63
Rate, % (SE)	100 (0)	94.0 (2.6)	84.5 (3.9)	79.8 (4.4)	75.0 (4.7)
Risk score 2					
Participants at risk, n	66	49	41	38	32
Rate, % (SE)	100 (0)	74.2 (5.4)	62.1 (6.0)	57.6 (6.1)	48.5 (6.2)
Risk score 3					
Participants at risk, n	20	12	8	3	2
Rate, % (SE)	100 (0)	60.0 (11.0)	40.0 (11.0)	15.0 (8.0)	10.0 (6.7)

Discussion

Summary

This study demonstrated the association and predictive value of malnutrition, inflammation, and stroke factors in the 2-year survival of octogenarians or nonagenarians after treatment with EVT for LEAD. Simple prognostic stratification using malnutrition, inflammation, and stroke factors will help clinicians in deciding the appropriate treatment.

For symptomatic LEAD necessitating revascularization, 2-year longevity is a fundamental consideration in deciding the treatment option. Surgical bypass, recommended by the recent guidelines for patients surviving for more than 2 years [3,4], was underused in older patients [23,24] because of the high perioperative morbidity, complications, and poor long-term outcomes in these patients [5-7]. Although EVT was reported to be a safe procedure for older patients with LEAD [5,6,13] and older patients may benefit from endovascular revascularization, the factors affecting 2-year longevity remain uncertain.

Association and Determining Variables for 2-Year Longevity

The ARA and grouped matrix method revealed that malnutrition and inflammation have meaningful roles in the 2-year longevity of octogenarians with atherosclerotic LEAD. The univariates affecting 2-year longevity were CHF, prior stroke, dialysis, ambulatory status, CLTI, BMI, total cholesterol, NLR, PLR, SII, CRP level, albumin, CONUT score, and GNRI. Those factors are clinically meaningful and consistent with prior reports by domain experts [25-28]. After adjusting for the covariates, NLR, prior stroke, and malnutrition (GNRI or CONUT score) remained independent predictors of 2-year longevity.

Impact of Immune-Inflammation Factors on Survival

Several immune-inflammatory factors (NLR, PLR, and SII) have been reported as prognostic markers in patients with LEAD in various clinical settings [29-31]. Previous studies have reported that an NLR >5 increases 1- and 2-year mortality rates, and it has been used as a major component of the Valladolid Critical Limb Ischaemia Risk Scale (ERICVA) model to predict 1-year amputation-free survival after revascularization for CLTI [32-34]. In patients with claudication, the cut-off value for major adverse cardiac events and disease severity was around 3.05 to 3.3 [35,36]. Our cohort study of 232 patients incorporated 40 (17.2%) patients with claudication, and the cut-off value for NLR was 3.89.

Elevated NLR reflects both the neutrophilia of inflammation (mediated by arachidonic acid metabolites and platelet-aggregating factors, cytotoxic oxygen-derived free radicals, and hydrolytic enzymes) [37] and relative lymphopenia, suggesting a deeper imbalance in the immunologic response, an increased expression of T helper 17 over regulatory T cells, and the activation of the interleukin-17 axis, which in turn is associated with vascular dysfunction, the progression of atherosclerosis, and vascular events [38,39]. In our study, we found that NLR had a higher predictive performance level than PLR and SII. Although a previous report showed that NLR increases with age [40], a high NLR played an additional role in assessing 2-year longevity in older patients with LEAD.

Nutrition

Impaired nutritional status, a functional disorder of frailty, increases the morbidity and mortality rates in older people [41]. Two objective indices of nutritional status, the GNRI and the CONUT score, have been reported as prognostic factors in patients with CLTI after EVT treatment [17,42]. It was found that patients with moderate to severe malnutrition (GNRI of <91-92 or CONUT of ≥4) had significantly higher cardiovascular and limb events compared with patients with normal nutritional status. The GNRI cut-off value of 90.3 was slightly lower in our study, which reflects a lower BMI in older patients with malnutrition.

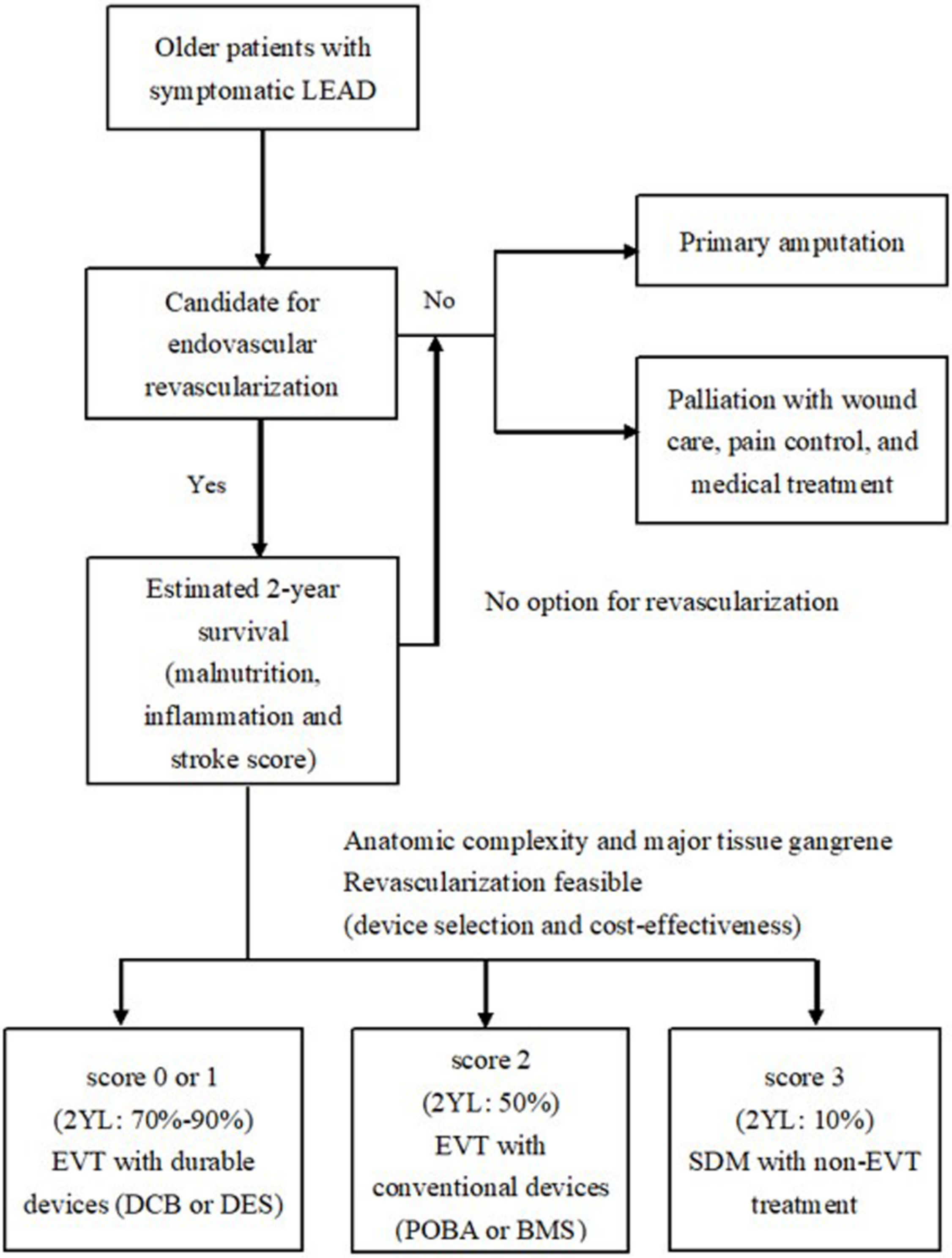
Chronic limb ischemia increases cytokine release, oxidative stress, and inflammatory cell accumulation, leading to malnutrition by affecting appetite, wasting resting energy, and increasing muscle protein breakdown [43,44]. Furthermore, malnutrition aggravates the progression of atherosclerosis. This cycle is called malnutrition-inflammation-atherosclerosis syndrome [44]. In this study, C statistics further validated the prognostic implication of these factors for 2-year survival, indicating the crucial role of malnutrition and impaired immune defenses in older people with preexisting atherosclerotic disease.

Comparisons With Prior Work

Moxey et al [25] compared the BASIL survival prediction model to the Finland National Vascular registry and Prevention of Infrainguinal Vein Graft Failure III (PREVENT III) models. The power of each model to predict mortality was evaluated by comparing the AUC for each ROC curve. The AUCs for 2-year mortality ranged from 0.533 to 0.664 in those models, which indicates weak to good prediction. In our study, the combination of independent covariates using C statistics further increased the AUC and determined the predictive value for 2-year longevity estimation.

Effective revascularization is the cornerstone treatment for patients with symptomatic LEAD, and EVT is the first-line treatment strategy in older people due to lower procedural risk. To aid the clinical decision making in daily practice, we used the malnutrition, inflammation, and stroke scores for risk stratification. Anatomical complexity and wound status, which are limb outcome predictors, help health care professionals select appropriate devices to treat limb ischemia and optimize the EVT results. Figure 7 illustrates the framework of the treatment decision. For patients without any or with only 1 malnutrition, inflammation, and stroke factor, 2-year longevity is around 70% to 90%. Thus, EVT should be performed and durable drug-coated devices should be applied to treat limb ischemia because of the longer life expectancy. For patients with 2 malnutrition, inflammation, and stroke factors, 2-year longevity is around 50%. EVT using standard balloon angioplasty or bare-metal stents might be cost-effective, as long-term vascular patency is not the focus for patients with an average life expectancy.

Figure 7. Treatment framework using the malnutrition, inflammation, and stroke score for older patients with symptomatic lower extremity arterial disease. 2YL: 2-year longevity; BMS: bare-metal stent; DCB: drug-coated balloon; DES: drug-eluting stent; EVT: endovascular therapy; LEAD: lower extremity arterial disease; POBA: plain old balloon angioplasty; SDM: shared decision making.



In contrast, 2-year longevity is only 10% in patients with 3 malnutrition, inflammation, and stroke factors. These patients may be appropriately treated with primary amputation or non-EVT treatment, such as spinal cord stimulation, lumbar sympathectomy, or intermittent pneumatic compression. Health care professionals can share their decision making with patients, families, and caregivers, who should have access to appropriate expertise when dealing with these challenging scenarios. This approach balances life expectancy, invasiveness, and benefits, which is consistent with the PLAN (patient risk estimation, limb staging, and anatomic pattern of arterial disease) concept recommended by current practical guidelines [45].

Study Limitations

This study had several limitations. First, this was an observational cohort study using a prospective database and all patients were treated in a single institution, indicating a potential selection bias. Second, the patient enrollment period spanned more than 10 years, and the outcomes and risk models may have changed over time due to new techniques, management, or reimbursement policies. Third, this study lacks data on subsequent changes in nutritional status. Therefore, we cannot determine whether nutrition status change had any effect on the clinical outcome of patients with peripheral arterial disease. Fourth, patients with poor cognitive function and uncooperative patients with dementia were not enrolled, which might have affected the results. We also did not enroll patients undergoing surgical revascularization, and outcome differences could not be determined between EVT and bypass surgery using the same scoring techniques. Despite the useful information regarding malnutrition, inflammation, and stoke scores in clinical practice, global assessment and shared decision making plays a crucial role in judging the balance between risk and effectiveness when dealing with challenging scenarios. Large-scale trials to prove the benefit of malnutrition, information, and stroke scores are warranted. Finally, ARA data mining techniques cannot rank association rules or determine their weight, and no measures of the importance of items in a rule or rule set have been proposed thus far [46].

Conclusions

Malnutrition, inflammation, and stroke factors are associated with 2-year longevity and play a crucial role in assessing older patients with LEAD. The presence of each malnutrition, inflammation, and stroke factor or their combination worsens the survival potential of these patients. Awareness of the likelihood of intermediate-term survival better informs the discussion with patients about the benefits of EVT, its attendant risks, and intermediate-term outcomes.

Acknowledgments

The authors appreciate the medical staff of the cardiac catheterization laboratory and the study coordinators who participated in this study.

Authors' Contributions

HLH and YLK conceived of and designed the study. HHC, SJJ, and CAH managed the acquisition of the data and the analysis and interpretation of the data. The drafting of the manuscript was performed with equal contributions from JYJ and IST. HLH, IST, and YLK critically revised the manuscript for important intellectual content. HLH gave final approval of the submitted version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Receiver-operating-characteristic curves of inflammatory variables for 2-year survival prediction.
[PNG File , 2476 KB - jmir_v22i12e17487_app1.png]

Multimedia Appendix 2
Supplementary Table 1. Patient demographics.
[PDF File (Adobe PDF File), 32 KB - jmir_v22i12e17487_app2.pdf]

Multimedia Appendix 3
Supplementary Table 2. Lesion and interventional procedure characteristics.
[PDF File (Adobe PDF File), 51 KB - jmir_v22i12e17487_app3.pdf]

Multimedia Appendix 4
Top 10 association rules of clinical data.
[DOCX File , 16 KB - jmir_v22i12e17487_app4.docx]

References

1. Fowkes FGR, Rudan D, Rudan I, Aboyans V, Denenberg JO, McDermott MM, et al. Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis. Lancet 2013 Oct 19;382(9901):1329-1340. [doi: 10.1016/S0140-6736(13)61249-0] [Medline: 23915883]
2. Diehm C, Schuster A, Allenberg JR, Darius H, Haberl R, Lange S, et al. High prevalence of peripheral arterial disease and co-morbidity in 6880 primary care patients: cross-sectional study. Atherosclerosis 2004 Jan;172(1):95-105. [doi: 10.1016/s0021-9150(03)00204-1] [Medline: 14709362]

3. Adam DJ, Beard JD, Cleveland T, Bell J, Bradbury AW, Forbes JF, BASIL trial participants. Bypass versus angioplasty in severe ischaemia of the leg (BASIL): multicentre, randomised controlled trial. Lancet 2005 Dec 03;366(9501):1925-1934. [doi: 10.1016/S0140-6736(05)67704-5] [Medline: 16325694]
4. Gerhard-Herman MD, Gornik HL, Barrett C, Barshes NR, Corriere MA, Drachman DE, et al. 2016 AHA/ACC Guideline on the Management of Patients With Lower Extremity Peripheral Artery Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2017 Mar 21;69(11):1465-1508 [FREE Full text] [doi: 10.1016/j.jacc.2016.11.008] [Medline: 27851991]
5. Brosi P, Dick F, Do DD, Schmidli J, Baumgartner I, Diehm N. Revascularization for chronic critical lower limb ischemia in octogenarians is worthwhile. J Vasc Surg 2007 Dec;46(6):1198-1207. [doi: 10.1080/17521740701702115] [Medline: 181514996]
6. Dosluoglu HH, Lall P, Cherr GS, Harris LM, Dryjski ML. Superior limb salvage with endovascular therapy in octogenarians with critical limb ischemia. J Vasc Surg 2009 Aug;50(2):305-15, 316.e1 [FREE Full text] [doi: 10.1016/j.jvs.2009.01.004] [Medline: 19631865]
7. Dick P, Barth B, Mlekusch W, Sabeti S, Amighi J, Schlager O, et al. Complications after peripheral vascular interventions in octogenarians. J Endovasc Ther 2008 Aug;15(4):383-389. [doi: 10.1583/08-2459.1] [Medline: 18729563]
8. Huang H, Chou H, Wu T, Chang S, Tsai Y, Hung S, et al. Endovascular intervention in Taiwanese patients with critical limb ischemia: patient outcomes in 333 consecutive limb procedures with a 3-year follow-up. J Formos Med Assoc 2014 Oct;113(10):688-695 [FREE Full text] [doi: 10.1016/j.jfma.2012.10.022] [Medline: 25240302]
9. Huang H, Tzeng I, Chou H, Hsieh C, Jang S, Ko Y, et al. Contemporary cardiovascular outcomes in Taiwanese patients undergoing endovascular therapy for symptomatic lower extremity peripheral arterial disease. J Formos Med Assoc 2019 Jun;119(6):1052-1060 [FREE Full text] [doi: 10.1016/j.jfma.2019.10.011] [Medline: 31672435]
10. Sun K, Bai F. Mining Weighted Association Rules without Preassigned Weights. IEEE Trans Knowl Data Eng 2008 Apr;20(4):489-495 [FREE Full text] [doi: 10.1109/tkde.2007.190723]
11. Datta S, Bose S. Mining and ranking association rules in support, confidence, correlation, and dissociation framework. 2015 Presented at: 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA) 2015; Nov 16-18, 2015; Durgapur, India URL: https://doi.org/10.1007/978-81-322-2695-6_13 [doi: 10.1007/978-81-322-2695-6_13]
12. Li Q, Zhang Y, Kang H, Xin Y, Shi C. Mining association rules between stroke risk factors based on the Apriori algorithm. Technol Health Care 2017 Jul 20;25(S1):197-205. [doi: 10.3233/THC-171322] [Medline: 28582907]
13. Huang H, Jimmy Juang J, Chou H, Hsieh C, Jang S, Cheng S, et al. Immediate results and long-term cardiovascular outcomes of endovascular therapy in octogenarians and nonoctogenarians with peripheral arterial diseases. Clin Interv Aging 2016;11:535-543 [FREE Full text] [doi: 10.2147/CIA.S106119] [Medline: 27217735]
14. Norgren L, Hiatt WR, Dormandy JA, Nehler MR, Harris KA, Fowkes FGR, TASC II Working Group. Inter-Society Consensus for the Management of Peripheral Arterial Disease (TASC II). J Vasc Surg 2007 Jan;45(Suppl S):S5-67 [FREE Full text] [doi: 10.1016/j.jvs.2006.12.037] [Medline: 17223489]
15. Narumi T, Arimoto T, Funayama A, Kadowaki S, Otaki Y, Nishiyama S, et al. Prognostic importance of objective nutritional indexes in patients with chronic heart failure. J Cardiol 2013 Nov;62(5):307-313 [FREE Full text] [doi: 10.1016/j.jjcc.2013.05.007] [Medline: 23806549]
16. Ignacio de Ulíbarri J, González-Madroño A, de Villar NGP, González P, González B, Mancha A, et al. CONUT: a tool for controlling nutritional status. First validation in a hospital population. Nutr Hosp 2005;20(1):38-45. [Medline: 15762418]
17. Yokoyama M, Watanabe T, Otaki Y, Watanabe K, Toshima T, Sugai T, et al. Impact of Objective Malnutrition Status on the Clinical Outcomes in Patients With Peripheral Artery Disease Following Endovascular Therapy. Circ J 2018 Feb 23;82(3):847-856 [FREE Full text] [doi: 10.1253/circj.CJ-17-0731] [Medline: 29187666]
18. Bouillanne O, Morineau G, Dupont C, Coulombel I, Vincent J, Nicolis I, et al. Geriatric Nutritional Risk Index: a new index for evaluating at-risk elderly medical patients. Am J Clin Nutr 2005 Oct;82(4):777-783. [doi: 10.1093/ajcn/82.4.777] [Medline: 16210706]
19. Tang J, Chuang L, Hsi E, Lin Y, Yang C, Chang H. Identifying the association rules between clinicopathologic factors and higher survival performance in operation-centric oral cancer patients using the Apriori algorithm. Biomed Res Int 2013;2013:359634 [FREE Full text] [doi: 10.1155/2013/359634] [Medline: 23984353]
20. Yang H, Xie Y, Ni J, Liu Y, Song R, Chen C, et al. Association Rule Analysis for Validating Interrelationships of Combined Medication of Compound Kushen Injection in Treating Colon Carcinoma: A Hospital Information System-Based Real-World Study. Evid Based Complement Alternat Med 2018;2018:4579801 [FREE Full text] [doi: 10.1155/2018/4579801] [Medline: 30228827]
21. Huang H, Juang JJ, Hsieh C, Chou H, Jang S, Ko Y. Risk stratification for low extremity amputation in critical limb ischemia patients who have undergone endovascular revascularization: A survival tree analysis. Medicine (Baltimore) 2019 Aug;98(33):e16809 [FREE Full text] [doi: 10.1097/MD.0000000000016809] [Medline: 31415395]
22. Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. 1994 Presented at: 20th International Conference on Very Large Data Bases (VLDB 1994); Sep 12-15, 1994; Santiago de Chile, Chile URL: http://www.vldb.org/conf/1994/P487.PDF

23. Reinecke H, Unrath M, Freisinger E, Bunzemeier H, Meyborg M, Lüders F, et al. Peripheral arterial disease and critical limb ischaemia: still poor outcomes and lack of guideline adherence. Eur Heart J 2015 Apr 14;36(15):932-938. [doi: 10.1093/eurheartj/ehv006] [Medline: 25650396]
24. Goodney PP, Travis LL, Nallamothu BK, Holman K, Suckow B, Henke PK, et al. Variation in the use of lower extremity vascular procedures for critical limb ischemia. Circ Cardiovasc Qual Outcomes 2012 Jan;5(1):94-102 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.111.962233] [Medline: 22147886]
25. Moxey PW, Brownrigg J, Kumar SS, Crate G, Holt PJ, Thompson MM, et al. The BASIL survival prediction model in patients with peripheral arterial disease undergoing revascularization in a university hospital setting and comparison with the FINNVASC and modified PREVENT scores. J Vasc Surg 2013 Jan;57(1):1-7 [FREE Full text] [doi: 10.1016/j.jvs.2012.04.074] [Medline: 23040797]
26. Soga Y, Iida O, Takahara M, Takahaera M, Hirano K, Suzuki K, et al. Two-year life expectancy in patients with critical limb ischemia. JACC Cardiovasc Interv 2014 Dec;7(12):1444-1449 [FREE Full text] [doi: 10.1016/j.jcin.2014.06.018] [Medline: 25523536]
27. Bradbury AW, Adam DJ, Bell J, Forbes JF, Fowkes FGR, Gillespie I, BASIL Trial Participants. Bypass versus Angioplasty in Severe Ischaemia of the Leg (BASIL) trial: A survival prediction model to facilitate clinical decision making. J Vasc Surg 2010 May;51(5)(Suppl):52S-68S [FREE Full text] [doi: 10.1016/j.jvs.2010.01.077] [Medline: 20435262]
28. Simons JP, Schanzer A, Flahive JM, Osborne NH, Mills JL, Bradbury AW, et al. Survival prediction in patients with chronic limb-threatening ischemia who undergo infrainguinal revascularization. J Vasc Surg 2019 Jun;69(6S):137S-151S.e3. [doi: 10.1016/j.jvs.2018.08.169] [Medline: 30497863]
29. Spark JI, Sarveswaran J, Blest N, Charalabidis P, Asthana S. An elevated neutrophil-lymphocyte ratio independently predicts mortality in chronic critical limb ischemia. J Vasc Surg 2010 Sep;52(3):632-636 [FREE Full text] [doi: 10.1016/j.jvs.2010.03.067] [Medline: 20573475]
30. Gary T, Pichler M, Belaj K, Hafner F, Gerger A, Froehlich H, et al. Platelet-to-lymphocyte ratio: a novel marker for critical limb ischemia in peripheral arterial occlusive disease patients. PLoS One 2013;8(7):e67688 [FREE Full text] [doi: 10.1371/journal.pone.0067688] [Medline: 23844064]
31. Tong Y, Tan J, Zhou X, Song Y, Song Y. Systemic immune-inflammation index predicting chemoradiation resistance and poor outcome in patients with stage III non-small cell lung cancer. J Transl Med 2017 Oct 31;15(1):221 [FREE Full text] [doi: 10.1186/s12967-017-1326-1] [Medline: 29089030]
32. Chan C, Puckridge P, Ullah S, Delaney C, Spark JI. Neutrophil-lymphocyte ratio as a prognostic marker of outcome in infrapopliteal percutaneous interventions for critical limb ischemia. J Vasc Surg 2014 Sep;60(3):661-668 [FREE Full text] [doi: 10.1016/j.jvs.2014.03.277] [Medline: 24816510]
33. Bhutta H, Agha R, Wong J, Tang TY, Wilson YG, Walsh SR. Neutrophil-lymphocyte ratio predicts medium-term survival following elective major vascular surgery: a cross-sectional study. Vasc Endovascular Surg 2011 Apr;45(3):227-231. [doi: 10.1177/1538574410396590] [Medline: 21289130]
34. Brizuela Sanz JA, González Fajardo JA, Taylor JH, Río Solá L, Muñoz Moreno MF, Vaquero Puerta C. Design of a New Risk Score in Critical Limb Ischaemia: The ERICVA Model. Eur J Vasc Endovasc Surg 2016 Jan;51(1):90-99 [FREE Full text] [doi: 10.1016/j.ejvs.2015.09.025] [Medline: 26602223]
35. Haumer M, Amighi J, Exner M, Mlekusch W, Sabeti S, Schlager O, et al. Association of neutrophils and future cardiovascular events in patients with peripheral artery disease. J Vasc Surgery 2005 Apr;41(4):610-617 [FREE Full text] [doi: 10.1016/j.jvs.2005.01.013] [Medline: 15874924]
36. Aykan A, Hatem E, Kalaycıoğlu E, Karabay CY, Zehir R, Gökdeniz T, et al. Neutrophil-to-lymphocyte ratio may be a marker of peripheral artery disease complexity. Anatol J Cardiol 2016 Jul;16(7):497-503 [FREE Full text] [doi: 10.5152/AnatolJCardiol.2015.6240] [Medline: 27004700]
37. Tamhane UU, Aneja S, Montgomery D, Rogers E, Eagle KA, Gurm HS. Association between admission neutrophil to lymphocyte ratio and outcomes in patients with acute coronary syndrome. Am J Cardiol 2008 Sep 15;102(6):653-657. [doi: 10.1016/j.amjcard.2008.05.006] [Medline: 18773982]
38. Cheng X, Yu X, Ding Y, Fu Q, Xie J, Tang T, et al. The Th17/Treg imbalance in patients with acute coronary syndrome. Clin Immunol 2008 Apr;127(1):89-97. [doi: 10.1016/j.clim.2008.01.009] [Medline: 18294918]
39. Madhur MS, Lob HE, McCann LA, Iwakura Y, Blinder Y, Guzik TJ, et al. Interleukin 17 promotes angiotensin II-induced hypertension and vascular dysfunction. Hypertension 2010 Feb;55(2):500-507 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.109.145094] [Medline: 20038749]
40. Davis JL, Moutinho V, Panageas KS, Coit DG. A peripheral blood biomarker estimates probability of survival: the neutrophil-lymphocyte ratio in noncancer patients. Biomark Med 2016 Sep;10(9):953-957 [FREE Full text] [doi: 10.2217/bmm-2016-0103] [Medline: 27537355]
41. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]

42. Shiraki T, Iida O, Takahara M, Masuda M, Okamoto S, Ishihara T, et al. The Geriatric Nutritional Risk Index is Independently Associated with Prognosis in Patients with Critical Limb Ischemia Following Endovascular Therapy. Eur J Vasc Endovasc Surg 2016 Aug;52(2):218-224 [FREE Full text] [doi: 10.1016/j.ejvs.2016.05.016] [Medline: 27357968]
43. Stenvinkel P, Heimbürger O, Lindholm B, Kaysen GA, Bergström J. Are there two types of malnutrition in chronic renal failure? Evidence for relationships between malnutrition, inflammation and atherosclerosis (MIA syndrome). Nephrol Dial Transplant 2000 Jul;15(7):953-960. [doi: 10.1093/ndt/15.7.953] [Medline: 10862630]
44. Pecoits-Filho R, Lindholm B, Stenvinkel P. The malnutrition, inflammation, and atherosclerosis (MIA) syndrome -- the heart of the matter. Nephrol Dial Transplant 2002;17(Suppl 11):28-31. [doi: 10.1093/ndt/17.suppl_11.28] [Medline: 12386254]
45. Conte MS, Bradbury AW, Kolh P, White JV, Dick F, Fitridge R, GVG Writing Group for the Joint Guidelines of the Society for Vascular Surgery (SVS), European Society for Vascular Surgery (ESVS), World Federation of Vascular Societies (WFVS). Global Vascular Guidelines on the Management of Chronic Limb-Threatening Ischemia. Eur J Vasc Endovasc Surg 2019 Jul;58(1S):S1-S109.e33. [doi: 10.1016/j.ejvs.2019.05.006] [Medline: 31182334]
46. Ceddia G, Martino LN, Parodi A, Secchi P, Campaner S, Masseroli M. Association rule mining to identify transcription factor interactions in genomic regions. Bioinformatics 2020 Feb 15;36(4):1007-1013. [doi: 10.1093/bioinformatics/btz687] [Medline: 31504203]

Abbreviations

ARA: association rules analysis
AUC: area under the curve
BASIL: Bypass Versus Angioplasty in Severe Ischaemia of the Leg
CBC: complete blood count
CHF: congestive heart failure
CLTI: chronic limb-threatening ischemia
CONUT: Controlling Nutritional Status
CRP: C-reactive protein
CVA: cerebrovascular accident
ERICVA: Valladolid Critical Limb Ischaemia Risk Scale
EVT: endovascular therapy
GNRI: Geriatric Nutritional Risk Index
HR: hazard ratio
LEAD: lower extremity arterial disease
NLR: neutrophil-lymphocyte ratio
PLAN: patient risk estimation, limb staging, and anatomic pattern of arterial disease
PLR: platelet-lymphocyte ratio
PREVENT III: Prevention of Infrainguinal Vein Graft Failure III
ROC: receiver operating characteristic
SII: systemic immune-inflammation index
WBC: white blood cell

Edited by G Eysenbach; submitted 18.12.19; peer-reviewed by D Newman, J Fan; comments to author 20.04.20; revised version received 19.05.20; accepted 11.11.20; published 01.12.20.

Please cite as:
Jhang JY, Tzeng IS, Chou HH, Jang SJ, Hsieh CA, Ko YL, Huang HL
Association Rule Mining and Prognostic Stratification of 2-Year Longevity in Octogenarians Undergoing Endovascular Therapy for Lower Extremity Arterial Disease: Observational Cohort Study
J Med Internet Res 2020;22(12):e17487
URL: https://www.jmir.org/2020/12/e17487
doi:10.2196/17487
PMID:33177036

Original Paper

Perceived Electronic Health Record Usability as a Predictor of Task Load and Burnout Among US Physicians: Mediation Analysis

Edward R Melnick[1], MD, MHS; Elizabeth Harry[2], MD; Christine A Sinsky[3], MD; Liselotte N Dyrbye[4], MD, MHPE; Hanhan Wang[5], MPS; Mickey Todd Trockel[6], MD, PhD; Colin P West[7,8], MD, PhD; Tait Shanafelt[5], MD

[1]Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT, United States
[2]University of Colorado School of Medicine, Aurora, CO, United States
[3]Professional Satisfaction and Practice Sustainability, American Medical Association, Chicago, IL, United States
[4]Department of Medicine Physician Well-Being Program, Mayo Clinic, Rochester, MN, United States
[5]Department of Medicine, Stanford School of Medicine, Palo Alto, CA, United States
[6]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, CA, United States
[7]Department of Internal Medicine, Mayo Clinic, Rochester, MN, United States
[8]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Edward R Melnick, MD, MHS
Department of Emergency Medicine
Yale University School of Medicine
464 Congress Ave, Suite 260
New Haven, CT, 06519
United States
Phone: 1 (203) 785 5174
Email: edward.melnick@yale.edu

Abstract

Background: Electronic health record (EHR) usability and physician task load both contribute to physician professional burnout. The association between perceived EHR usability and workload has not previously been studied at a national level. Better understanding these interactions could give further information as to the drivers of extraneous task load.

Objective: This study aimed to determine the relationship between physician-perceived EHR usability and workload by specialty and evaluate for associations with professional burnout.

Methods: A secondary analysis of a cross-sectional survey of US physicians from all specialties was conducted from October 2017 to March 2018. Among the 1250 physicians invited to respond to the subsurvey analyzed here, 848 (67.8%) completed it. EHR usability was assessed with the System Usability Scale (SUS; range: 0-100). Provider task load (PTL) was assessed using the mental demand, physical demand, temporal demand, and effort required subscales of the National Aeronautics and Space Administration Task Load Index (range: 0-400). Burnout was measured using the Maslach Burnout Inventory.

Results: The mean scores were 46.1 (SD 22.1) for SUS and 262.5 (SD 71.7) for PTL. On multivariable analysis adjusting for age, gender, relationship status, medical specialty, practice setting, hours worked per week, and number of nights on call per week, physician-rated EHR usability was associated with PTL, with each 1-point increase in SUS score (indicating more favorable) associated with a 0.57-point decrease in PTL score ($P<.001$). On mediation analysis, higher SUS score was associated with lower PTL score, which was associated with lower odds of burnout.

Conclusions: A strong association was observed between EHR usability and workload among US physicians, with more favorable usability associated with less workload. Both outcomes were associated with the odds of burnout, with task load acting as a mediator between EHR usability and burnout. Improving EHR usability while decreasing task load has the potential to allow practicing physicians more working memory for medical decision making and patient communication.

(J Med Internet Res 2020;22(12):e23382) doi:10.2196/23382

KEYWORDS

electronic health record; EHR; usability; cognitive load; System Usability Scale; physician task load; NASA Task Load Index; physician; burnout; stress

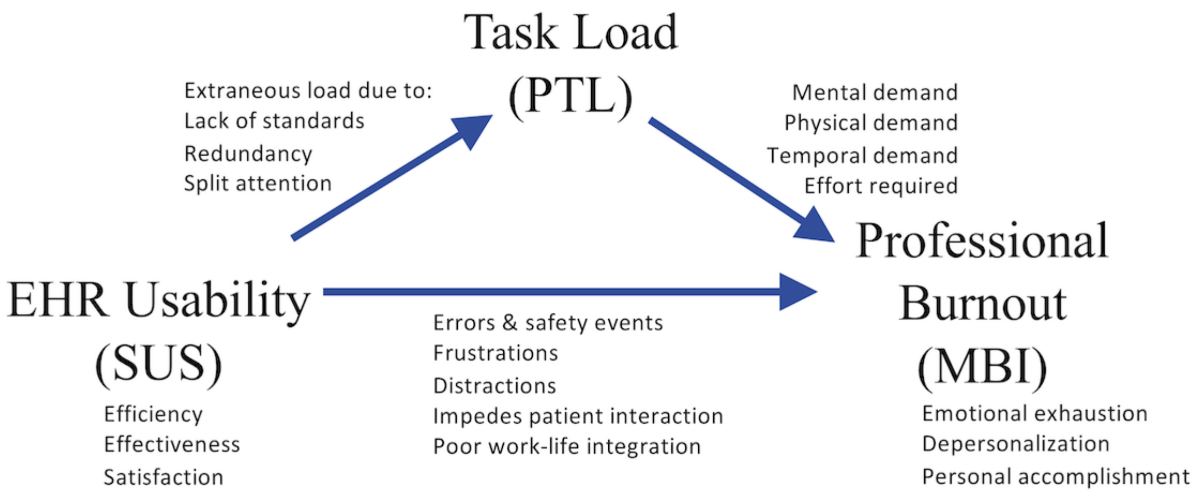
Introduction

Occupational burnout in physicians is a complex phenomenon with multiple interdependent drivers related to practice efficiency, organizational culture, and personal resilience [1]. In the domain of practice efficiency, there is growing evidence supporting an association between electronic health record (EHR) usability and physician burnout [2-7]. Usability is the extent to which technology can be used effectively, efficiently, and satisfactorily based on its design and integration into a specific context of use [8]. Although the EHR has been lauded as a solution to improve health care quality and safety, there is increasing evidence that EHR usability can cause harm [9-12]. Current EHR usability challenges have resulted in systems that many find unnecessarily complex and prone to error, thereby increasing physician cognitive load and resulting in technological errors that can sometimes reach the patient [10-13]. Partially due to excessive time spent on EHR activities, EHR usability is a specific source of physician dissatisfaction and stress [7,14-16]. The topic has received increasing attention since a 2017 systematic review demonstrated a paucity of published studies and standardized reporting on EHR usability evaluation [17-20]. A recently published cross-sectional national survey of physicians from all specialty disciplines identified a strong association between higher physician-perceived EHR usability (assessed by the industry standard, the System Usability Scale [SUS]) and lower odds of physician burnout [7].

In addition to usability, the clerical burden associated with documentation, order entry, inbox management, and other EHR administrative tasks (eg, prior authorizations, documentation of care consistent with quality measures) that are not necessarily intrinsic to the practice of medicine also contributes to excessive time spent on the EHR, as well as unnecessary cognitive burden [16,21,22]. Administrative tasks such as these are independent of EHR design—indeed, the EHR may offer advantages by providing a systematic structure and record to such tasks [21]. According to cognitive load theory, tasks like these can overwhelm limited working memory, the process our minds use to input and respond to all information [23]. Cognitive load refers to the amount of working memory used and comprises three components: intrinsic (complexity of the task itself), extraneous (how the task is presented), and germane (the workload of learning the task or content) load [24,25]. When users are overloaded, data is “shed” or lost, which puts users at risk of committing an error [26]. Using the mental demand, physical demand, temporal demand, and effort required subscales of the National Aeronautics and Space Administration Task Load Index (NASA-TLX), a national survey of 4622 physicians conducted by our team identified a strong association between provider task load (PTL) and burnout but did not assess EHR usability [27,28]. Two studies found enhanced EHR usability is associated with reduced physician cognitive load in simulated EHR environments [29,30].

We propose a conceptual framework in which task load acts as a mediator between EHR usability and professional burnout (Figure 1). For example, a poor EHR interface with patient lab values may present redundant data on the screen, creating unnecessary extraneous load that increases the mental demand of a physician’s task. Ongoing exposure to this interface could contribute to burnout for some physicians. Independent of this task, the poor usability could impede patient communication, thereby diminishing the physician’s sense of purpose in their work, thus leading to burnout [10,31].

Figure 1. Conceptual framework with provider task load (PTL) as a mediator between electronic health record (EHR) usability and professional burnout. MBI: Maslach Burnout Inventory; SUS: System Usability Scale.



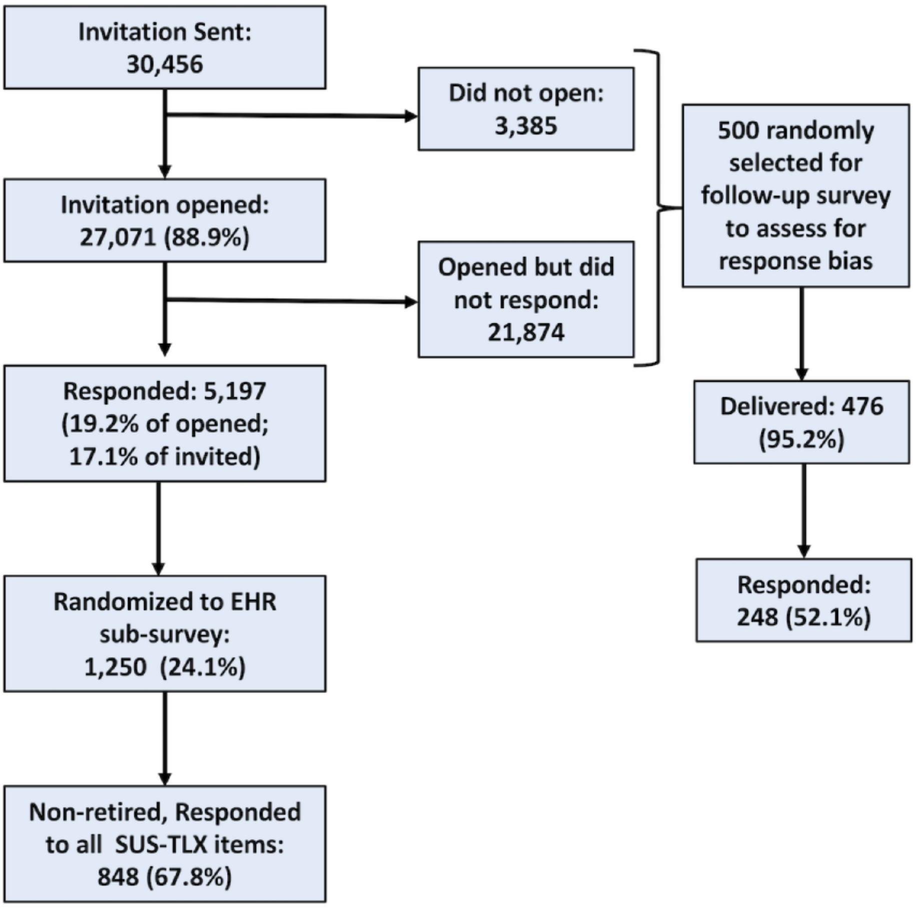
To our knowledge, the association between perceived EHR usability and workload has not previously been studied at a national level. We conducted a cross-sectional analysis of a national survey to determine the association between physician-perceived EHR usability (using the SUS) and the workload of a clinical workday stratified by specialty and practice setting (using the mental demand, physical demand, temporal demand, and effort required subscales of the NASA-TLX, henceforth referred to as PTL). We hypothesized that greater EHR usability scores (as measured by the SUS) would correlate with lower PTL scores. Also, given that both the SUS and PTL have been shown to relate to physician burnout, we hypothesized that they would both remain associated with burnout after adjusting for personal and professional characteristics. Better understanding these interactions could give further information as to the drivers of extraneous task load.

Methods

Study Design, Setting, Participants, and Data Collection

A secondary analysis of a cross-sectional wellness study of US physicians from all subspecialties was performed. The original survey collected data between October 12, 2017, and March 15, 2018. The sample was assembled using the American Medical Association Physician Masterfile, a nearly complete record of all US physicians, independent of American Medical Association membership. Participation involved voluntary completion of an anonymous electronic survey. Full details of the sampling strategy, recruitment, data collection, and assessment for response bias have previously been reported [32]. Briefly, 30,456 physicians from the Masterfile were invited to participate (Figure 2). Of these, 5197 (17.1%) physicians participated in the study [32]. A random group of 1250 of the participants was invited to complete a subsurvey evaluating their EHR's usability. Responders who were retired from clinical practice were excluded from the analysis. To evaluate for response bias, an intensive follow-up survey was conducted in a sample of nonresponders. There were 248 (52.1% of 476 invited) participants in the follow-up survey. The Stanford University and Mayo Clinic institutional review boards reviewed and approved the study protocol.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram of subject enrollment, inclusion criteria, and exclusion criteria. EHR: electronic health record; SUS-TLX: System Usability Scale-Task Load Index.



Study Measures

Study measures included items pertaining to participants' demographic characteristics (age, gender, relationship status), medical specialty, hours worked per week, number of nights on call per week, practice setting, symptoms of burnout, and perception of their EHR's usability and daily clinical workload.

EHR Usability

Physician-perceived EHR usability was measured using the SUS, an industry standard for a quick, reliable measurement of technology usability [33-35]. The SUS includes 10 items, each on a 5-point Likert scale from Strongly Disagree to Strongly Agree that is scored from 0 to 4, with each even-numbered question reverse coded. The items are summed and then multiplied by 2.5 to normalize scores from 0 to 100, with higher scores indicating higher usability. Consistent with convention [34,36], the language of the SUS was modified such that references to "the system" being evaluated were changed to "my EHR."

Provider Task Load

Provider task load was measured using 4 items (the mental demand, physical demand, temporal demand, and effort required subscales) from the NASA-TLX [28]. The rationale for exclusion of the frustration and performance NASA-TLX items from our measurement of PTL is that on principal component analysis with oblimin rotation and Kaiser normalization to determine underlying patterns between the 6 items of the NASA-TLX and the emotional exhaustion and depersonalization scales of the Maslach Burnout Inventory, frustration and performance clustered as one component along with emotional exhaustion and depersonalization scores, suggesting that these domains of the NASA-TLX are measures of work-related distress and would be expected to be collinear with burnout measures. Inclusion of the mental demand, physical demand, temporal demand, and effort required NASA-TLX items is consistent with previous reports assessing workload of physicians and nurses [27,37]. Respondents were prompted to "reflect on a day [they] performed clinical work during the last 1-2 weeks that is representative of a typical current clinical work day" and rate their perception of each subscale demand type on a scale of 0 to 100 (with 100 being the highest level of demand) [38,39]. The 4 scores were summed for a composite score ranging from 0 to 400 [40].

Burnout

Burnout was measured using the validated criterion standard, the Maslach Burnout Inventory [41-44]. Respondents with a high score on the emotional exhaustion (≥27) or depersonalization subscale (≥10) were considered to have at least 1 symptom of burnout [41,45-47].

Statistical Analysis

Standard descriptive statistics were used to characterize the physician sample that responded to the EHR usability subsurvey and the PTL items of the survey. Associations between variables were evaluated using the Wilcoxon rank sum test (continuous variables) or χ^2 test (categorical variables), as appropriate. Univariable linear regression was used to examine the association between EHR usability and PTL. On preliminary analysis controlling for medical specialty, specialties with smaller numbers of participants in the subsurvey had considerable variability. To control for this variation, specialties with fewer than 20 participants were grouped together in a pooled category of "Other" specialties. The Other category included these specialties (number of respondents in parentheses): neurosurgery (9), ophthalmology (8), otolaryngology (10), other (18), physical medicine and rehabilitation (15), preventive medicine & occupational medicine (4), radiation oncology (5), and urology (7). Two multivariable linear regression analyses were performed to investigate whether the relationship between SUS score and burnout is mediated by PTL. Demographic and professional factors included in the multivariable regression analyses were age, gender, relationship status, hours worked per week, medical specialty, nights on call, and practice setting. All tests were 2-sided, with a type I error level of .05. Analyses were completed using R statistical software (R Foundation for Statistical Computing, Vienna, Austria, Version 3.6.0) with the exception of the mediation analyses, which were conducted using the PROCESS version 3 macro for SPSS (IBM Corporation) [48].

Results

Participants

There were 5197 responders to the full survey. A randomly selected group of 1250 of these responders received the EHR usability subsurvey. Among these responders, the 848 individuals (67.8%) who responded to all SUS and PTL items were included in the present analysis. The demographic characteristics (age, gender, and medical specialty) of the respondents included in this analysis were generally similar to the full survey respondents and US physicians (Table 1).

Table 1. Demographic characteristics of survey responders.

Characteristic	EHR usability subsurvey responders (N=848)
Gender, n (%)	
Male	493 (58.1)
Female	348 (41.0)
Other	1 (0.1)
Missing	6 (0.7)
Age (years), median (IQR)	53.0 (42.0, 61.0)
Age (years), n (%)	
<35	61 (7.2)
35-44	200 (23.6)
45-54	183 (21.6)
55-64	259 (30.5)
>/=65	123 (14.5)
Missing	22 (2.6)
Specialty, n (%)	
Anesthesiology	36 (4.2)
Dermatology	23 (2.7)
Emergency medicine	54 (6.4)
Family medicine	55 (6.5)
Radiology	37 (4.4)
Neurology	32 (3.8)
Obstetrics and gynecology	42 (5.0)
Pathology	25 (2.9)
Psychiatry	51 (6.0)
Other	76 (9.0)
General internal medicine	77 (9.1)
Internal medicine subspecialty	113 (13.3)
General pediatrics	46 (5.4)
Pediatric subspecialty	53 (6.2)
General surgery	33 (3.9)
General surgery subspecialty	58 (6.8)
Orthopedic surgery	35 (4.1)
Missing	2 (0.2)
Hours worked per week, median (IQR)	50.0 (40.0, 60.0)
Hours worked per week, n (%)	
<40 h	126 (14.9)
40-49 h	171 (20.2)
50-59 h	215 (25.4)
60-69 h	187 (22.1)
70-79 h	76 (9.0)
>80 h	70 (8.3)
Missing	3 (0.4)
Nights on call per week, median (IQR)	1.0 (0.0, 2.0)

Characteristic	EHR usability subsurvey responders (N=848)
Primary practice setting, n (%)	
Private practice	388 (45.8)
Academic medical center	274 (32.3)
Veterans hospital	16 (1.9)
Active military practice	9 (1.1)
Other	160 (18.9)
Missing	1 (0.1)
Relationship status, n (%)	
Single	105 (12.4)
Married	687 (81.0)
Partnered	39 (4.6)
Widowed/widower	8 (0.9)
Missing	9 (1.1)

SUS and PTL Scores

Physician-perceived EHR usability and PTL scores were similar to those previously reported in the primary analyses of this survey [7,27]. For this analysis, the mean SUS score was 46.1 (SD 22.1; range: 0-100; IQR 30-62.5), and the mean composite PTL score was 262.5 (SD 71.7; range: 0-400; IQR 215-315) with mean subscale scores of 70.9 (SD 20.8) for mental demand, 47.8 (SD 27.3) for physical demand, 68.3 (SD 24.9) for time demand, and 75.6 (SD 21.1) for effort required. On univariate analysis, each 1-point increase in SUS score (indicating greater EHR usability) was associated with a 0.55-point decrease in PTL score (*P*<.001; Figure 3). The relationship between perceived EHR usability and composite PTL scores is shown for the 17 specialty discipline categories in Figure 4.

Figure 3. Scatter plot of provider task load (PTL) sum scores (range: 0-400) by System Usability Scale (SUS) scores (range: 0-100) with regression line.

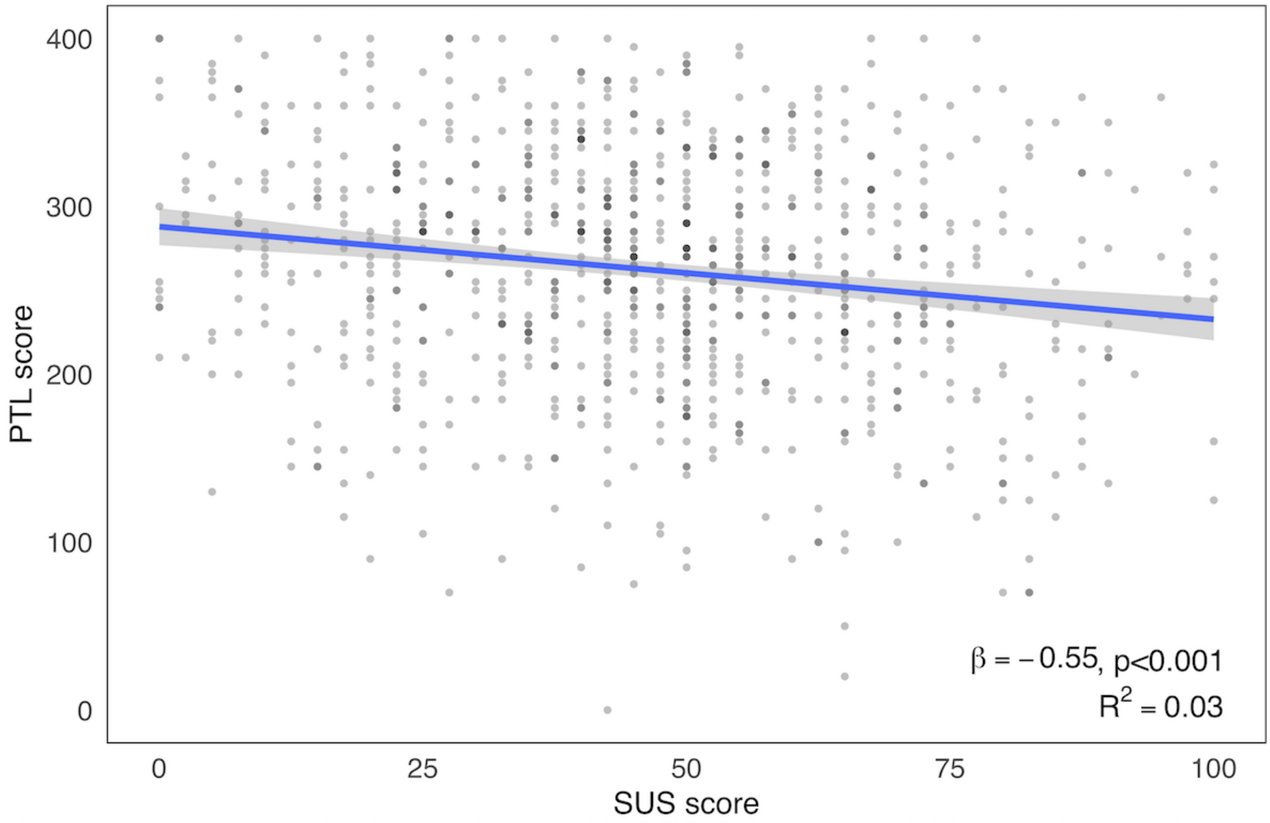
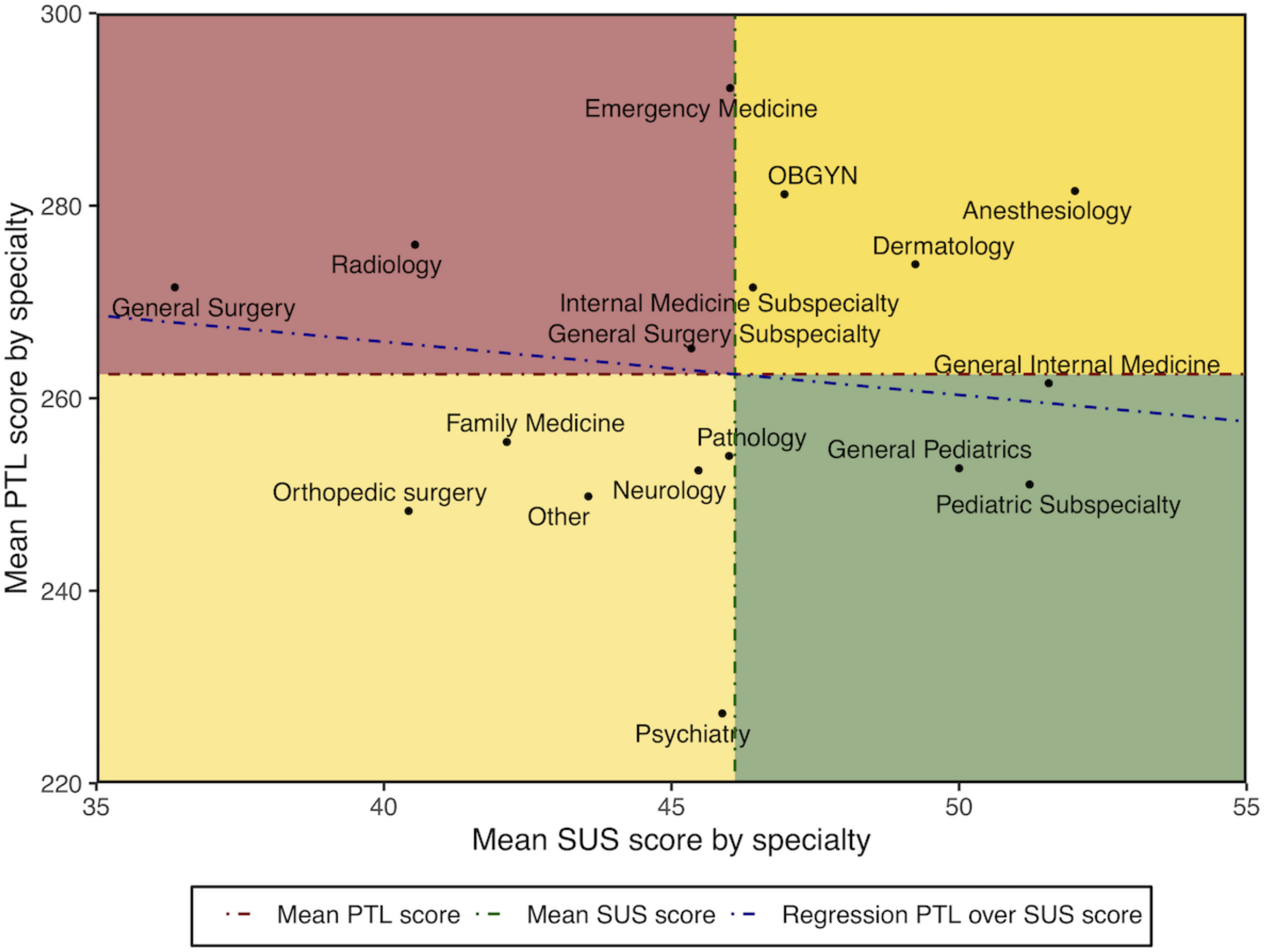


Figure 4. Scatter plot of provider task load (PTL) as assessed by the modified NASA-TLX and electronic health record usability as assessed by the System Usability Scale (SUS) by specialty with regression line. Note that higher SUS indicates more favorable usability, whereas higher PTL indicates increased task load (less favorable).



Multivariable and Mediation Analyses

On multivariable analysis adjusting for age, gender, relationship status, medical specialty, practice setting, hours worked per week, and number of nights on call per week, EHR SUS scores were associated with PTL. Each 1-point increase in SUS score was associated with a 0.57-point decrease in composite PTL score (*P*<.001; Table 2). Gender; age; practicing emergency medicine, anesthesiology, or psychiatry; hours worked per week; and number of nights on call per week were also associated with composite PTL scores in this model. Being female, practicing emergency medicine or anesthesiology, more hours worked per week, and more nights on call per week were all associated with higher PTL as a measure of workload, whereas being older and practicing psychiatry were both associated with lower workload.

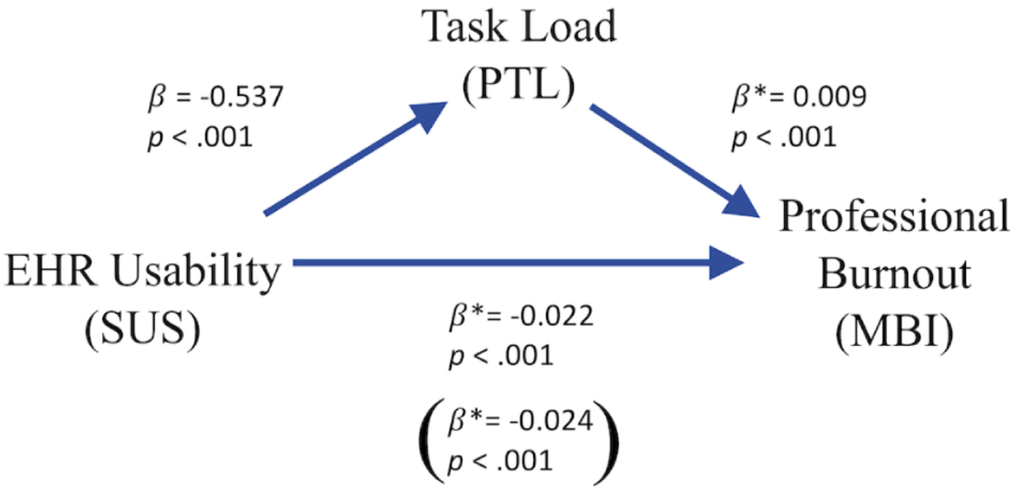
Table 2. Predictors of provider task load in multivariable linear regression models among practicing physicians in 2017.

Predictor	Coefficient (SE)	*P* value
SUS[a] score	−0.57 (0.03)	.001
Gender (reference: male)		
Female	12.59 (1.68)	.03
Missing/other	19.26 (10.24)	.58
Age, for each year older	−0.45 (0.07)	.048
Specialty (reference: general internal medicine)		
Anesthesiology	30.6 (4.18)	.03
Dermatology	13.68 (5.03)	.42
Emergency medicine	47.04 (3.83)	<.001
Family medicine	−2.87 (3.71)	.82
Radiology	13.33 (4.25)	.36
Neurology	−1.18 (4.4)	.94
Obstetrics and gynecology	9.1 (4)	.5
Pathology	−4.13 (4.83)	.8
Psychiatry	−29.11 (3.84)	.03
Other	−9.66 (3.42)	.4
Internal medicine subspecialty	8.83 (3.13)	.41
General pediatrics	−0.76 (3.99)	.96
Pediatric subspecialty	−14.01 (3.84)	.28
General surgery	−1.23 (4.4)	.93
General surgery subspecialty	−6.76 (3.71)	.59
Orthopedic surgery	−9.47 (4.24)	.51
Hours worked per week, for each additional hour	0.98 (0.05)	<.001
Nights on call per week, for each additional call	3.26 (0.37)	.01
Primary practice setting (reference: private practice)		
Academic medical center	−0.62 (1.77)	.92
Veterans hospital	2.22 (5.22)	.9
Active military practice	−31.44 (6.9)	.18
Other	4.5 (2.07)	.52
Relationship status (reference: single)		
Married	14.12 (2.24)	.06
Partnered	19.95 (3.93)	.13
Widowed/widower	46.33 (8.61)	.11

[a] SUS: System Usability Scale.

In the first ordinary least squares regression model of the mediation analysis (Figure 5), higher SUS was significantly related to lower PTL scores (β=−.537, 95% CI −0.755 to −0.319; SE 0.111; *P*<.001). In the second logistic regression model, which included SUS and PTL as predictor variables of burnout, both SUS and PTL were significantly associated with burnout (OR 0.978, 95% CI 0.972 to 0.985 and OR 1.009, 95% CI 1.007 to 1.011, respectively). The bootstrap confidence intervals derived from 5000 samples indicated that the indirect effect of PTL on the association between SUS and burnout was significant (effect=−0.005, 95% CI −0.007 to −0.003). From this result, higher SUS was associated with lower PTL, which was associated with lower odds of burnout.

Figure 5. Mediation analysis results showing the quantitative relationships between provider task load (PTL) as a mediator between electronic health record (EHR) usability and professional burnout. The values for EHR usability to professional burnout display the association between the System Usability Scale (SUS) and Maslach Burnout Inventory (MBI) after controlling for PTL and, parenthetically, the association between SUS and MBI when PTL is not statistically controlled for. *log-odds, since this is a logistic regression.



Discussion

Principal Results

The findings of this national study indicate that physician-perceived EHR usability scores are strongly associated with physician daily task load. After adjusting for multiple personal and professional characteristics, individual physicians' composite PTL scores were 0.55 points lower for each 1 point more favorable EHR usability, as measured by the SUS. The mediation analysis supports the proposed conceptual framework in which task load acts as a mediator between EHR usability and professional burnout. Despite the strong statistically significant associations found, the amount of variability in PTL as a measure of workload explained by EHR usability was small (R^2=0.03). This finding indicates that factors other than EHR usability appear to be the primary drivers of physician workload as assessed by the mental demand, physical demand, temporal demand, and effort required subscales of the NASA-TLX. In the primary analysis of this survey, age, gender, medical specialty, hours worked per week, number of nights on call per week, and practice setting were all associated with PTL as a measure of workload [27]. To put these findings in context, a 10-point lower PTL was associated with 30% lower odds of burnout in the primary analysis [27]. An 18-point higher average SUS score (associated with a 10-point lower PTL in this analysis) would give the EHR a grade of D (instead of F) and be above the score of Microsoft Excel (Microsoft Corporation, 2009) [7].

Exploring this relationship and variation in EHR usability and PTL as a measure of workload by specialty reveals that some specialties with higher task load rated their EHRs more favorably (eg, anesthesiology and dermatology) while other specialties associated with lower task load (eg, orthopedic surgery and family medicine) rated their EHRs less favorably. This finding suggests the relationship between EHR usability and task load may not be due to physicians in specialties with higher task load rating their EHR less favorably. For example, anesthesiologists may benefit from the EHR's ability to provide certain information rapidly in the operating room, yet still have higher task load due to factors in the anesthesiology clinical context that are not EHR-related. However, only three specialties (emergency medicine, anesthesiology, and psychiatry) remained predictive of task load after controlling for EHR usability, gender, age, hours worked per week, and number of nights on call per week. The analysis is likely underpowered to explain specific differences between individual specialties. On balance, SUS appears to have only a small influence on task load, although the effect of SUS on PTL varies by specialty.

Comparison With Prior Work

It is also notable that in this analysis, lower EHR usability and higher workload were both associated with the odds of burnout after controlling for multiple personal and professional characteristics, suggesting that these are distinct domains that both represent potential improvement targets to reduce physician burnout. A study of 46 participants' perception of the usability and task load of three popular websites found no association between usability and task load as assessed by the SUS and NASA-TLX [49]. Important differences between that study and the present study were that SUS scores were only in the high range (compared to the EHR) and that they measured both domains for individual tasks as opposed to in aggregate as we have here. The primary analysis of this national survey of 5197 physicians found that age, gender, relationship status, hours worked per week, and practicing certain medical specialties were all associated with the odds of burnout [32]. In this analysis, gender was not significantly predictive of the odds of burnout. Given this finding and the growing literature on the association of gender on physician burnout [32,50,51], future work should explore the interactions between these variables and further evaluate how EHR usability and task load vary by personal and professional characteristics. For example, a mixed-methods assessment of 25 intensive care physicians running simulated cases found gender-based differences in perceived EHR workload stress, satisfaction, and usability as assessed by the NASA-TLX and SUS [52]. Modest redesign of computerized reminders was associated with improved usability and decreased mental workload for 16 nurses in a simulated environment as assessed by the NASA-TLX [53]. A randomized crossover trial of 7 pediatric surgeons reported improved SUS scores and lower cognitive workload scores when order sets were systematically developed [54]. Future work could explore the relationship between EHR usability, specific task load, and professional burnout across and between different medical specialty disciplines and EHR vendor products as well as further differentiate the contribution of administrative burden to these areas by specialty. Policy makers should also explore the potential savings to the health care system that might be realized by improving EHR usability and how removing required administrative tasks that are not intrinsic to the practice of medicine could decrease task load and burnout.

Limitations

This study is subject to several limitations. First, although the association between SUS and PTL scores is statistically significant, cross-sectional data cannot prove causation or the potential direction of effect. Second, with all survey research, response bias and representativeness of the study sample are potential limitations. An assessment for response bias was employed in this study, including evaluation of a random sample of nonresponders to the initial survey who participated in an incentivized follow-up survey. Although women were more likely to participate in the EHR usability subsurvey, participants were generally representative of US physicians with respect to age, years in practice, and prevalence of burnout. Third, self-reported data is subject to limitations [55,56]. Physician perceptions of usability and workload may not accurately reflect reality [20]. For example, one dimension of usability is error tolerance, a system's ability to prevent or recover from an error; due to poor usability, EHR users may not be aware of errors they make in the EHR [11,12]. In addition, individual respondents could have a tendency to provide similar responses to different questions, which could contribute to a common method bias [57]. Fourth, although the SUS and NASA-TLX are considered the industry standards to assess technology usability and task load, they are intended to assess a specific task within a single system, not an entire class of software or an entire day of clinical work. Instead, this analysis provides a global assessment of EHR usability and provider task load and, therefore, cannot account for differences between specific tasks or vendors' products and may be subject to recall bias. Fifth, physician respondents could conflate EHR usability issues with regulatory and clerical demands that manifest in the EHR but may be unrelated to EHR user interface design [21,22]. If that is the case, the association of clerical burden with task load and burnout could be stronger than that of EHR usability [7,16].

Despite these limitations, as the first national study exploring the relationship between EHR usability and workload in practice across medical specialties and practice settings, this analysis adds an important dimension to existing knowledge about factors associated with PTL and physician burnout. Our findings are consistent with multiple smaller studies of task load in both simulated [29,30] and real world settings [58,59] suggesting that more usable EHR interfaces are associated with lower cognitive load. Technology changes physicians' information gathering and reasoning strategies, thereby shaping cognitive behavior [60]. Cognitive load theory should be taken into consideration to design EHR interfaces and workflows that meet users' needs and are free from unnecessary extraneous cognitive load, such that the EHR becomes part of a physician work environment more closely resembling a “manageable cockpit” that is streamlined, ergonomic, and safe [10,61,62].

Conclusion

As assessed by US physicians using standardized metrics of technology and workload, a strong association was observed between EHR usability and task load, with more favorable usability associated with lower task load. Although EHR usability was significantly related to workload, the effect size was small, indicating that factors other than the EHR appear to be the primary drivers of workload. Both outcomes were also associated with the odds of burnout. Efforts to address physician burnout should attend to both improving EHR usability and addressing other drivers of task load, particularly extraneous load. Improving EHR usability while decreasing extraneous task load has the potential to allow practicing physicians more available working memory for medical decision making and patient communication. Specific areas to target could include consolidating the display of related information, reducing redundancy of information, and increasing standardization [63,64].

Acknowledgments

The authors would like to thank Bidisha Nath, MD, MPH, for her contribution to the literature search and manuscript coordination. Funding for this study was provided by the Stanford Medicine WellMD Center, the American Medical Association, and the Mayo Clinic Department of Medicine Program on Physician Well-Being. ERM is supported in part by two American Medical Association Practice Transformation Initiatives and the National Institute on Drug Abuse of the National Institutes of Health under Award Number UH3DA047003. The content is solely the responsibility of the authors and does not necessarily represent the official views of the American Medical Association or the National Institutes of Health.

Authors' Contributions

ERM, EH, CAS, and TS significantly contributed to the conception and design of the study. TS, LND, CAS, MTT, and CPW acquired the data. ERM, EH, HW, and TS analyzed the data. ERM drafted the initial manuscript. All authors were involved in data interpretation, revised the manuscript, and approved the final version submitted for publication. HW and TS had access to the data in the study and take responsibility for data integrity and accuracy. ERM takes responsibility for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Bohman B, Dyrbye L, Sinsky C, Linzer M, Olson K, Babbott S, et al. Physician well-being: the reciprocity of practice efficiency, culture of wellness, and personal resilience.: NEJM Catalyst; 2017. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.17.0429 [accessed 2020-12-07]
2. Olson K, Sinsky C, Rinne S, Long T, Vender R, Mukherjee S, et al. Cross-sectional survey of workplace stressors associated with physician burnout measured by the Mini-Z and the Maslach Burnout Inventory. Stress Health 2019 Apr;35(2):157-175. [doi: 10.1002/smi.2849] [Medline: 30467949]
3. Gardner RL, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, et al. Physician stress and burnout: the impact of health information technology. J Am Med Inform Assoc 2019 Feb 01;26(2):106-114. [doi: 10.1093/jamia/ocy145] [Medline: 30517663]
4. Tai-Seale M, Dillon EC, Yang Y, Nordgren R, Steinberg RL, Nauenberg T, et al. Physicians' Well-Being Linked To In-Basket Messages Generated By Algorithms In Electronic Health Records. Health Aff (Millwood) 2019 Jul;38(7):1073-1078. [doi: 10.1377/hlthaff.2018.05509] [Medline: 31260371]
5. Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, Grumbach K. Electronic health records and burnout: Time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. J Am Med Inform Assoc 2020 Apr 01;27(4):531-538. [doi: 10.1093/jamia/ocz220] [Medline: 32016375]
6. Micek MA, Arndt B, Tuan W, Trowbridge E, Dean SM, Lochner J, et al. Physician Burnout and Timing of Electronic Health Record Use. ACI Open 2020 Feb 06;04(01):e1-e8. [doi: 10.1055/s-0039-3401815]
7. Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The Association Between Perceived Electronic Health Record Usability and Professional Burnout Among US Physicians. Mayo Clin Proc 2020 Mar;95(3):476-487 [FREE Full text] [doi: 10.1016/j.mayocp.2019.09.024] [Medline: 31735343]
8. Bevan N, Carter J, Harker S. Human-Computer Interaction Design and Evaluation. ISO 9241-11 Revised: What Have We Learnt About Usability Since 1998?: Springer International Publishing; 2015:143-151.
9. Zahabi M, Kaber DB, Swangnetr M. Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation. Hum Factors 2015 Aug;57(5):805-834. [doi: 10.1177/0018720815576827] [Medline: 25850118]
10. Gellert G, Webster S, Gillean J, Melnick E, Kanzaria H. Should US doctors embrace electronic health records? BMJ 2017 Dec 24;356:j242. [Medline: 28119282]
11. Ratwani RM, Savage E, Will A, Fong A, Karavite D, Muthu N, et al. Identifying Electronic Health Record Usability And Safety Challenges In Pediatric Settings. Health Aff (Millwood) 2018 Dec;37(11):1752-1759. [doi: 10.1377/hlthaff.2018.0699] [Medline: 30395517]
12. Ratwani RM, Savage E, Will A, Arnold R, Khairat S, Miller K, et al. A usability and safety analysis of electronic health records: a multi-center study. J Am Med Inform Assoc 2018 Sep 01;25(9):1197-1201. [doi: 10.1093/jamia/ocy088] [Medline: 29982549]
13. Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic Health Record Usability Issues and Potential Contribution to Patient Harm. JAMA 2018 Dec 27;319(12):1276-1278 [FREE Full text] [doi: 10.1001/jama.2018.1171] [Medline: 29584833]
14. Friedberg MW, Chen PG, Van Busum KR, Aunon F, Pham C, Caloyeras J, et al. Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy. Rand Health Q 2014;3(4):1 [FREE Full text] [Medline: 28083306]
15. Babbott S, Manwell LB, Brown R, Montague E, Williams E, Schwartz M, et al. Electronic medical records and physician stress in primary care: results from the MEMO Study. J Am Med Inform Assoc 2014 Feb;21(e1):e100-e106 [FREE Full text] [doi: 10.1136/amiajnl-2013-001875] [Medline: 24005796]
16. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional Satisfaction. Mayo Clin Proc 2016 Jul;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]
17. Walji MF, Kalenderian E, Piotrowski M, Tran D, Kookal KK, Tokede O, et al. Are three methods better than one? A comparative assessment of usability evaluation methods in an EHR. Int J Med Inform 2014 May;83(5):361-367 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.010] [Medline: 24583045]
18. Ratwani R, Zachary Hettinger A, Kosydar A, Fairbanks R, Hodgkins M. A framework for evaluating electronic health record vendor user-centered design and usability testing processes. J Am Med Inform Assoc 2017 Apr 01;24(e1):e35-e39 [FREE Full text] [doi: 10.1093/jamia/ocw092] [Medline: 27375292]

19. Ratwani RM, Reider J, Singh H. A Decade of Health Information Technology Usability Challenges and the Path Forward. JAMA 2019 Feb 26;321(8):743-744. [doi: 10.1001/jama.2019.0161] [Medline: 30715133]
20. Ellsworth MA, Dziadzko M, O'Horo JC, Farrell AM, Zhang J, Herasevich V. An appraisal of published usability evaluations of electronic health records via systematic review. J Am Med Inform Assoc 2017 Jan;24(1):218-226. [doi: 10.1093/jamia/ocw046] [Medline: 27107451]
21. Downing NL, Bates DW, Longhurst CA. Physician Burnout in the Electronic Health Record Era: Are We Ignoring the Real Cause? Ann Intern Med 2018 Jul 03;169(1):50-51. [doi: 10.7326/M18-0139] [Medline: 29801050]
22. Tutty MA, Carlasare LE, Lloyd S, Sinsky CA. The complex case of EHRs: examining the factors impacting the EHR user experience. J Am Med Inform Assoc 2019 Jul 01;26(7):673-677 [FREE Full text] [doi: 10.1093/jamia/ocz021] [Medline: 30938754]
23. Sweller J. Cognitive Load Theory. In: Psychology of Learning and Motivation. San Diego, CA: Elsevier Academic Press; 2011:37-76.
24. Paas F, Renkl A, Sweller J. Cognitive Load Theory and Instructional Design: Recent Developments. Educational Psychologist 2003 Mar;38(1):1-4 [FREE Full text] [doi: 10.1207/s15326985ep3801_1]
25. DeLeeuw K, Mayer R. A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. Journal of Educational Psychology 2008;100(1):223-234 [FREE Full text] [doi: 10.1037/0022-0663.100.1.223]
26. Staal M. Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework. 2004. URL: https://ntrs.nasa.gov/search.jsp?R=20060017835 [accessed 2020-12-07]
27. Harry E, Sinsky C, Dyrbye LN, Makowski MS, Trockel M, Tutty M, et al. Physician Task Load and the Risk of Burnout Among US Physicians in a National Survey. Jt Comm J Qual Patient Saf 2020 Oct 04;S1553-7250(20):30246-30244. [doi: 10.1016/j.jcjq.2020.09.011] [Medline: 33168367]
28. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2016 Nov 05 Presented at: Human Factors and Ergonomics Society Annual Meeting; 2006; Washington, DC p. 904-908 URL: http://journals.sagepub.com/doi/abs/10.1177/154193120605000909 [doi: 10.1177/154193120605000909]
29. Mazur LM, Mosaly PR, Moore C, Marks L. Association of the Usability of Electronic Health Records With Cognitive Workload and Performance Levels Among Physicians. JAMA Netw Open 2019 Apr 05;2(4):e191709 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.1709] [Medline: 30951160]
30. Pollack AH, Pratt W. Association of Health Record Visualizations With Physicians' Cognitive Load When Prioritizing Hospitalized Patients. JAMA Netw Open 2020 Jan 03;3(1):e1919301 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.19301] [Medline: 31940040]
31. Flanagan ME, Militello LG, Rattray NA, Cottingham AH, Frankel RM. The Thrill Is Gone: Burdensome Electronic Documentation Takes Its Toll on Physicians' Time and Attention. J Gen Intern Med 2019 Jul;34(7):1096-1097 [FREE Full text] [doi: 10.1007/s11606-019-04898-8] [Medline: 31011960]
32. Shanafelt TD, West CP, Sinsky C, Trockel M, Tutty M, Satele DV, et al. Changes in Burnout and Satisfaction With Work-Life Integration in Physicians and the General US Working Population Between 2011 and 2017. Mayo Clin Proc 2019 Sep;94(9):1681-1694 [FREE Full text] [doi: 10.1016/j.mayocp.2018.10.023] [Medline: 30803733]
33. Brooke J. In: Jordan PW, Weerdmeester BA, McClelland IL, editors. SUS: A Quick and Dirty Usability Scale. London: Taylor and Francis; 1996.
34. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
35. Brooke J. SUS: a retrospective. Journal of Usability Studies 2013;8(2):29-40 [FREE Full text]
36. Lewis J, Sauro J. The Factor Structure of the System Usability Scale. In: Human Centered Design. Berlin, Heidelberg: Springer; 2009:94-103.
37. Tubbs-Cooley HL, Mara CA, Carle AC, Gurses AP. The NASA Task Load Index as a measure of overall workload among neonatal, paediatric and adult intensive care nurses. Intensive Crit Care Nurs 2018 Jun;46:64-69. [doi: 10.1016/j.iccn.2018.01.004] [Medline: 29449130]
38. Rutledge T, Stucky E, Dollarhide A, Shively M, Jain S, Wolfson T, et al. A real-time assessment of work stress in physicians and nurses. Health Psychol 2009 Mar;28(2):194-200. [doi: 10.1037/a0013145] [Medline: 19290711]
39. Weigl M, Müller A, Vincent C, Angerer P, Sevdalis N. The association of workflow interruptions and hospital doctors' workload: a prospective observational study. BMJ Qual Saf 2012 May;21(5):399-407. [doi: 10.1136/bmjqs-2011-000188] [Medline: 22190539]
40. Rölfing JD, Nørskov JK, Paltved C, Konge L, Andersen SAW. Failure affects subjective estimates of cognitive load through a negative carry-over effect in virtual reality simulation of hip fracture surgery. Adv Simul (Lond) 2019;4:26 [FREE Full text] [doi: 10.1186/s41077-019-0114-9] [Medline: 31832243]
41. Maslach C, Jackson S, Leiter M. Maslach Burnout Inventory Manual. Palo Alto, CA: Consulting Psychologists Press; 1997.

42. Rafferty JP, Lemkau JP, Purdy RR, Rudisill JR. Validity of the Maslach Burnout Inventory for family practice physicians. J Clin Psychol 1986 May;42(3):488-492. [doi: 10.1002/1097-4679(198605)42:3<488::aid-jclp2270420315>3.0.co;2-s] [Medline: 3711351]
43. Lee RT, Ashforth BE. A meta-analytic examination of the correlates of the three dimensions of job burnout. J Appl Psychol 1996 Apr;81(2):123-133. [doi: 10.1037/0021-9010.81.2.123] [Medline: 8603909]
44. Leiter M, Durup J. The discriminant validity of burnout and depression: A confirmatory factor analytic study. Anxiety, Stress, & Coping 2007 May 29;7(4):357-373. [doi: 10.1080/10615809408249357]
45. Shanafelt TD, Bradley KA, Wipf JE, Back AL. Burnout and self-reported patient care in an internal medicine residency program. Ann Intern Med 2002 Mar 05;136(5):358-367. [doi: 10.7326/0003-4819-136-5-200203050-00008] [Medline: 11874308]
46. Thomas NK. Resident burnout. JAMA 2004 Dec 15;292(23):2880-2889. [doi: 10.1001/jama.292.23.2880] [Medline: 15598920]
47. Rosen IM, Gimotty PA, Shea JA, Bellini LM. Evolution of sleep quantity, sleep deprivation, mood disturbances, empathy, and burnout among interns. Acad Med 2006 Jan;81(1):82-85. [doi: 10.1097/00001888-200601000-00020] [Medline: 16377826]
48. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis. In: A Regression-Based Approach Second Edition. New York: Guilford Press; 2017.
49. Longo L. Experienced mental workload, perception of usability, their interaction and impact on task performance. PLoS One 2018;13(8):e0199661 [FREE Full text] [doi: 10.1371/journal.pone.0199661] [Medline: 30067747]
50. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Arch Intern Med 2012 Oct 08;172(18):1377-1385. [doi: 10.1001/archinternmed.2012.3199] [Medline: 22911330]
51. Shanafelt TD, Hasan O, Dyrbye LN, Sinsky C, Satele D, Sloan J, et al. Changes in Burnout and Satisfaction With Work-Life Balance in Physicians and the General US Working Population Between 2011 and 2014. Mayo Clin Proc 2015 Dec;90(12):1600-1613. [doi: 10.1016/j.mayocp.2015.08.023] [Medline: 26653297]
52. Khairat S, Coleman C, Ottmar P, Bice T, Koppel R, Carson SS. Physicians' gender and their use of electronic health records: findings from a mixed-methods usability study. J Am Med Inform Assoc 2019 Dec 01;26(12):1505-1514 [FREE Full text] [doi: 10.1093/jamia/ocz126] [Medline: 31504578]
53. Saleem JJ, Patterson ES, Militello L, Anders S, Falciglia M, Wissman JA, et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. J Am Med Inform Assoc 2007;14(5):632-640 [FREE Full text] [doi: 10.1197/jamia.M2163] [Medline: 17600106]
54. Avansino J, Leu MG. Effects of CPOE on provider cognitive workload: a randomized crossover trial. Pediatrics 2012 Sep;130(3):e547-e552. [doi: 10.1542/peds.2011-3408] [Medline: 22891236]
55. Donaldson S, Grant-Vallone E. Understanding self-report bias in organizational behavior research. J Bus Psychol 2002;17(2):245-260. [doi: 10.1023/A:1019637632584]
56. Hunt M, Auriemma J, Cashaw ACA. Self-report bias and underreporting of depression on the BDI-II. J Pers Assess 2003 Feb;80(1):26-30. [doi: 10.1207/S15327752JPA8001_10] [Medline: 12584064]
57. Conway JM, Lance CE. What Reviewers Should Expect from Authors Regarding Common Method Bias in Organizational Research. J Bus Psychol 2010 May 22;25(3):325-334. [doi: 10.1007/s10869-010-9181-6]
58. Ariza F, Kalra D, Potts HW. How Do Clinical Information Systems Affect the Cognitive Demands of General Practitioners? Usability Study with a Focus on Cognitive Workload. J Innov Health Inform 2015 Nov 20;22(4):379-390 [FREE Full text] [doi: 10.14236/jhi.v22i4.85] [Medline: 26855272]
59. Agha Z. Quantifying electronic medical record usability to improve clinical workflow-final report (Prepared by the Veterans Medical Research Foundation under Grant No R01 HS021290). Rockville, MD: Agency for Healthcare Research and Quality; 2017.
60. Patel VL, Kushniruk AW, Yang S, Yale JF. Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning. J Am Med Inform Assoc 2000;7(6):569-585 [FREE Full text] [doi: 10.1136/jamia.2000.0070569] [Medline: 11062231]
61. Ray J, Ratwani R, Sinsky C, Frankel R, Friedberg M, Powsner S, et al. Six habits of highly successful health information technology: powerful strategies for design and implementation. J Am Med Inform Assoc 2019 Oct 01;26(10):1109-1114 [FREE Full text] [doi: 10.1093/jamia/ocz098] [Medline: 31265064]
62. Sinsky C, Privitera M. Creating a "Manageable Cockpit" for Clinicians: A Shared Responsibility. JAMA Intern Med 2018 Jun 01;178(6):741-742. [doi: 10.1001/jamainternmed.2018.0575] [Medline: 29582080]
63. de Jong T. Cognitive load theory, educational research, and instructional design: some food for thought. Instr Sci 2009 Aug 27;38(2):105-134. [doi: 10.1007/s11251-009-9110-0]
64. Sinsky CA, Rule A, Cohen G, Arndt BG, Shanafelt TD, Sharp CD, et al. Metrics for assessing physician activity using electronic health record log data. J Am Med Inform Assoc 2020 Apr 01;27(4):639-643 [FREE Full text] [doi: 10.1093/jamia/ocz223] [Medline: 32027360]

Abbreviations

EHR: electronic health record
NASA-TLX: National Aeronautics and Space Administration Task Load Index
PTL: provider task load
SUS: System Usability Scale

Edited by G Eysenbach, R Kukafka; submitted 12.08.20; peer-reviewed by S Rinne, R Hilliard, J Brooke, C Son; comments to author 02.09.20; revised version received 15.10.20; accepted 07.12.20; published 22.12.20.

Please cite as:
Melnick ER, Harry E, Sinsky CA, Dyrbye LN, Wang H, Trockel MT, West CP, Shanafelt T
Perceived Electronic Health Record Usability as a Predictor of Task Load and Burnout Among US Physicians: Mediation Analysis
J Med Internet Res 2020;22(12):e23382
URL: http://www.jmir.org/2020/12/e23382/
doi:10.2196/23382
PMID:33289493

Original Paper

Biological, Chemical, and Nutritional Food Risks and Food Safety Issues From Italian Online Information Sources: Web Monitoring, Content Analysis, and Data Visualization

Barbara Tiozzo[1], MAC; Mirko Ruzza[1], MAC; Valentina Rizzoli[1], PhD; Laura D'Este[2], BCS; Mosè Giaretta[1], MSc; Licia Ravarotto[1], MSc

[1]Department for Training, Communication and Support Services, Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro, Italy
[2]IT Service, Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro, Italy

Corresponding Author:
Barbara Tiozzo, MAC
Department for Training, Communication and Support Services
Istituto Zooprofilattico Sperimentale delle Venezie
viale dell'Università 10
Legnaro, 35020
Italy
Phone: 39 049 8084448
Email: btiozzo@izsvenezie.it

Abstract

Background: With rapid evolution of the internet and web 2.0 apps, online sources have become one of the main channels for most people to seek food risk information. Thus, it would be compelling to analyze the coverage of online information sources related to biological, chemical, and nutritional food risks, and related safety issues, to understand the type of content that online readers are exposed to, possibly influencing their perceptions.

Objective: The aim of this study was to identify the types of online sources that are predominantly covering this theme, and the topics that have received the most attention in terms of coverage and engagement on social media.

Methods: We performed an analysis of big data related to food risks by combining web monitoring techniques, content analysis, and data visualization of a large amount of unstructured text. Using a dictionary-based approach, a web monitoring app was instructed to automatically collect web content referring to the food risk and safety field. Data were retrieved from March 2017 to February 2018. The validated corpus (N=12,163) was subject to automatic and manual content analysis. Results were combined with descriptive statistics extracted from Web-Live and processed with Qlik Sense.

Results: Nutritional risks and news about outbreaks, controls, and alerts were the most widely covered topics. Thematic sources devoted major attention to nutritional topics, whereas national sources covered food risks, especially during food emergencies. Regarding engagement on social media, readers' interest was higher for nutritional topics and animal welfare. Although traditional sources still publish a great amount of content related to food risks and safety, new mediators have emerged as alternative sources for food risk information.

Conclusions: This mixed methodological approach was demonstrated to be a useful means for obtaining an accurate characterization of the online discourse on food risks, and can provide insight into how the monitored sources contribute to the process of risk communication.

(J Med Internet Res 2020;22(12):e23438) doi:10.2196/23438

KEYWORDS

big data; online information sources; web monitoring; content analysis; data visualization; food risks; risk communication

Introduction

Background

Risk perception and communication research has shown that many consumers worry about the quality and safety of the food they buy and eat [1-3]. To combat this concern, consumers actively engage in food risk information-seeking to gain useful recommendations about how to prevent and potentially face these risks [4].

Although television remains the most common source of information about food risks [3,5], online sources (ie, the internet and digital media) have become some of the main channels of searching for food risk information for most people [6-9]. The internet offers a wide variety of sources for health searches. In addition to news websites, many other forms of online health communities serve as online information sources [7,10,11], including search engines, health websites, social networking sites such as Facebook and Twitter, and social question and answer services, where risk information is permanently available and accessible [12].

Since information on food risks is widely sought online, it would be compelling to analyze the coverage of food risk by online information sources. According to previous literature, assessing the quantity of information about food risks that is available online can be of paramount importance to estimating how much information people may be exposed to and the extent to which people might engage with that information, possibly becoming an amplification station in terms of risk perception [13-15].

Web monitoring methodologies and techniques represent some of the best ways to analyze how much and where a topic is represented on the internet. Web monitoring is a set of specific activities that are planned to monitor and collect text from online sources to gain a better understanding of, and to measure the extent, time, and content that people are talking about with respect to a topic, person, brand, product, or service [16-19]. Monitoring what is being said online provides opportunities and benefits in terms of competitive analysis; sentiment and opinion analysis; knowledge discovery; consumer knowledge management; management and decision-making processes; social media strategies; creation or innovation of products and services; and predictions of scenarios, trends, and events [20].

As an application of text mining techniques [21,22], the web monitoring process basically consists of techniques for collecting, extracting, analyzing, and presenting online data that are relevant to the practitioner's research aims [23]. For example, using a dictionary-based approach [24,25], the internet environment is scanned to gather relevant text according to a predefined taxonomy that determines the content that has to be automatically collected using, for example, application programming interfaces and/or crawlers. The selected content is then extracted and analyzed to discover patterns or relationships, which can be translated into valuable information [26-28].

Although web monitoring and text mining are a strong tradition in corporate communication [16,18,19,29,30], their application is a novelty in the domain of food risk communication. To the best of our knowledge, this study represents a first attempt to evaluate the online coverage of food risks using a big data approach.

Study Aims

This study was designed to analyze big data related to food risks by combining web monitoring, content analysis, and data visualization approaches to determine the type of information Italian online sources provide about food risks and safety. Specifically, we addressed the following research questions:

RQ1: Which online information sources have covered food risks the most and received major engagement on social media?

RQ2: Which food risk topics have received major attention in terms of coverage and engagement on social media?

RQ3: Is there any difference in the coverage of food risk and safety topics among the monitored sources?

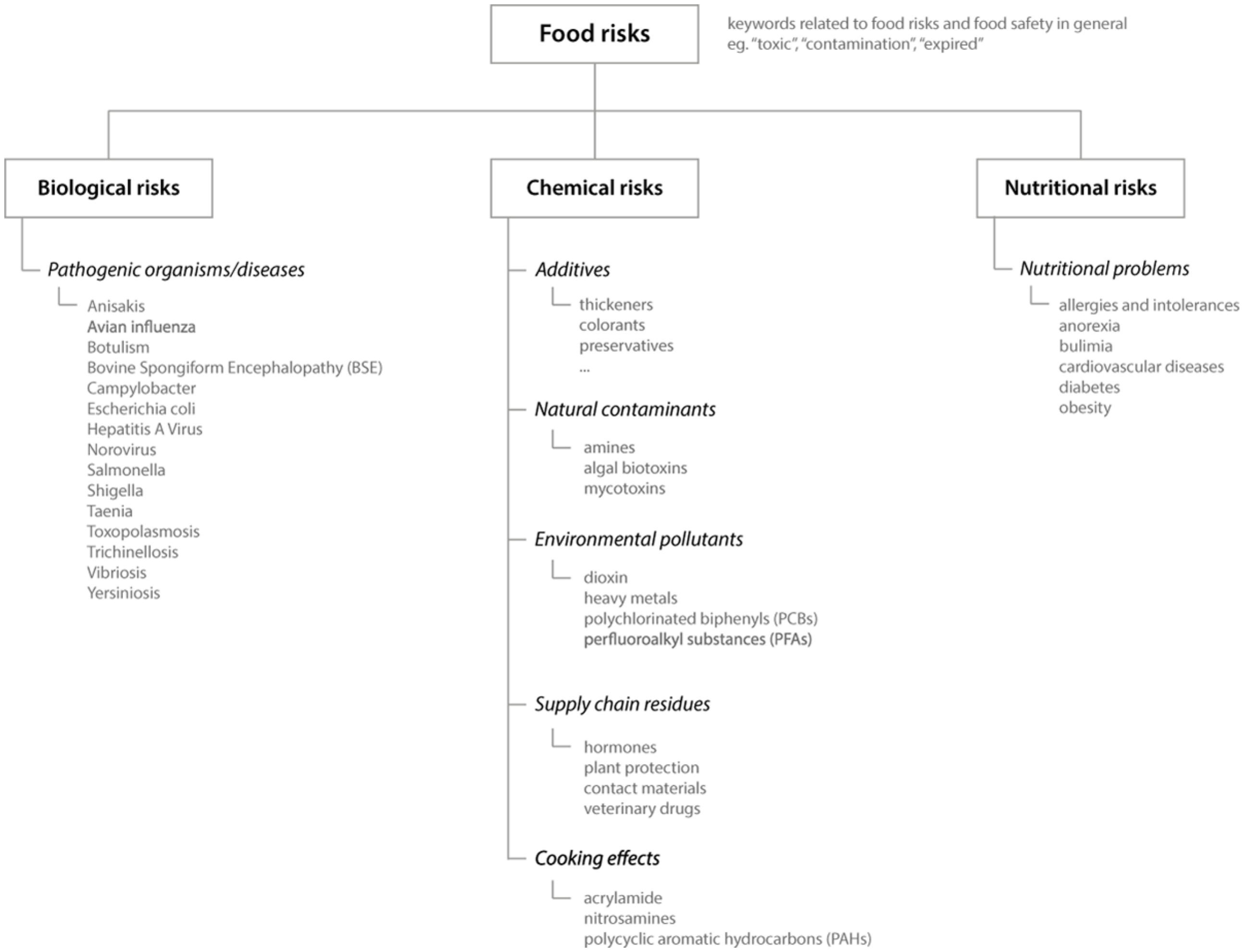
Based on these questions, four main topic categories were considered: biological, chemical, nutritional risks, and food safety issues. Biological risks include all hazards caused by bacteria, parasites, and viruses that can lead to food spoilage and possibly food poisoning for the consumer [31], such as *Salmonella* and *Campylobacter*. Chemical risks are hazards to human health that are derived from atoms, molecules, or substances, and can be present in raw ingredients for many reasons or can be formed in food during the production chain or the managing of food by the consumer after purchase [31-33]. A nutritional risk refers to the likelihood of adverse events occurring as a result of both poor nutritional quality and the amount of food consumed. This risk is associated with exposure to an inadequate diet, over or under in terms of quality and amount, for a certain period. Discourses on these risks belong to the broader concept of food safety that goes beyond the single risk agent and generally refers to all practices that are used to keep food safe, including cultural, scientific, and economical aspects, through the involvement of several different figures mandated to maintain food safety throughout the food chain.

Methods

Phase 1: Development of the Monitoring Profile

According to the logical scheme shown in Figure 1, a list of terms and rules referring to the four main categories under study was defined. Two of the authors, being senior experts with long-term experience in food risk management, identified the terms that were further validated by an external senior expert. The validated terms served as a starting point to define the monitoring profile, which is shown in Multimedia Appendix 1.

Figure 1. Logical scheme of the monitoring profile.



Phase 2: Data Acquisition (Web Monitoring)

Data were collected from March 1, 2017 to February 28, 2018 using Web-Live, a web monitoring app [22] (for further details, see Multimedia Appendix 2) that was instructed to work according to the monitoring profile. Every day throughout the reference period, the content automatically collected the day before was listed in descending order with respect to the number of estimated views. The list was read and analyzed to select the content characterized by greater visibility and higher relevance in terms of food risk significance. This selection and validation process was performed in Web-Live according to the following procedure: based on the reading of the title, snippet, or complete text of the content, content that was not relevant to the topic of food risks was excluded; only content assessed as relevant was included in the final corpus, up to a pre-established limit of 50 items per day; and each day, a preliminary tag was assigned to each piece of daily validated content to track the main topic it discussed and to inform further content analysis procedures (first level of tagging).

Phase 3: Data Analysis and Interpretation

After the monitoring phase, the content and statistics related to the validated corpus were extracted from Web-Live.

To answer RQ2, we first performed an automated content analysis. Automatic content analysis procedures based on text mining techniques have gained importance and popularity in the digital media environment due to the presence of larger datasets [34-36], and these methods have already successfully been used to analyze text that refers to food risks and safety issues [37-40]. A subcorpus composed of all validated content without content published on social media was extracted. Texts from social media were excluded because these types of texts differ in terms of length (they are normally shorter) and linguistic register. The resulting subcorpus was subject to a preprocessing phase [41] using TaLTaC2 software version 2.10.2 [42]. By means of the Reinert method, implemented in the R interface Iramuteq (version 0.7) [43], clusters of words referring to the same class of meaning (ie, lexical worlds) [44] were individuated to reveal patterns that characterize the online content on food risks. The associations among the clusters and the presence of the clusters over time were then observed using R software v. 3.5.1.

The results from automated content analysis guided the manual qualitative analysis of the retrieved content [45]. According to an open coding process [46], a label was assigned to each content in the corpus (second level of tagging). Starting from the lexical words identified by the automatic analysis and according to the tags that were preliminarily associated with the 50 pieces of daily validated content (first level of tagging), we proceeded with drawing up a list of the tags in use (Multimedia Appendix 3). When a new topic emerged, the list of tags was reviewed and refined iteratively. If two or more content items referred to the same topic, the content was assigned the same tag. In this manner, new tags were added to the list as they were created, thereby facilitating the researcher's task in assigning them without overlaps or duplication. Mutually exclusive tags were applied according to the prevalent topic treated in the content. These tags were then grouped into broader tags (third and fourth levels of tagging). The labeling process was performed separately by two trained coders using Microsoft Excel. Any discrepancies were discussed and resolved until an agreement was reached, and a third coder was involved in a supervisory role and guaranteed consistency in the tag assignment.

To answer RQ1, the sources with more than 10 content items were selected and subsequently grouped into a set of 6 source categories based on criteria such as editorial line, distribution on the territory, and type of author (Multimedia Appendix 4).

According to recent literature, data visualization is not only considered an output of the research process or a way of communicating results but is further considered to be an integral part of the research process itself, because one of its goals is to derive insight from massive, dynamic, ambiguous, and often conflicting data [47]. With this approach, statistics related to the validated corpus extracted from Web-Live were combined with topic and source categories resulting from the content analysis using the data visualization software Qlik Sense, which provided an overview of the distribution of content among the different sources and topics.

Results

Data Retrieval

The web monitoring phase returned 12,163 validated content items from the 209,847 raw data retrieved by the web monitoring app. These 12,163 validated items were considered the final corpus that was further analyzed. Overall, this content reached a total of 1,117,491 interactions. All of the text and the statistics related to the final analyzed sample are available in Multimedia Appendix 5 and in a dedicated web app [48].

Online Information Sources Related to Food Risk and Safety Topics (RQ1)

As shown in Table 1, the greatest amount of content from the total corpus (N=12,163) was published on web sources and the remaining content was published on social media. Overall, the content was published by 3230 unique sources that were equally active both as web sources and social media accounts. Among these, 240 unique sources published more than 10 content items (classified sources), whereas 2990 sources did not reach the threshold of 10 (nonclassified sources). Therefore, almost half of the total content was published by a larger number of different sources on an occasional basis, while the other half was published by a limited number of sources, mainly websites. The content published on social media was notably more extensive for the nonclassified sources.

Table 1. Major sources based on amount of content and website/social media ratio.

Source type	Unique sources (N=3230)			Contents (N=12,163)		
	Total, n (%)	Websites, n (%)	Social media, n (%)	Total, n (%)	Websites, n (%)	Social media, n (%)
Classified sources	240 (7.43)	207 (86.3)	33 (13.7)	6474 (53.23)	5823 (89.94)	651 (10.06)
Nonclassified sources	2990 (92.57)	1431 (47.86)	1559 (52.14)	5689 (46.8)	3200 (56.25)	2489 (43.75)
All sources	3230 (100.00)	1638 (50.71)	1592 (49.29)	12,163 (100.00)	9023 (74.18)	3140 (25.82)

A specialized website about food safety issues (ilfattoalimentare.it) emerged as the top source. Other major sources of interest were about the weather, the environment, and health news media (eg, meteoweb.eu, greenme.it, greenstyle.it) (Table 2).

Among the classified sources (Table 3), the thematic sources devoted major attention to the topic, followed by generalist news sources, local sources, and national sources.

Table 2. Major sources by coverage.

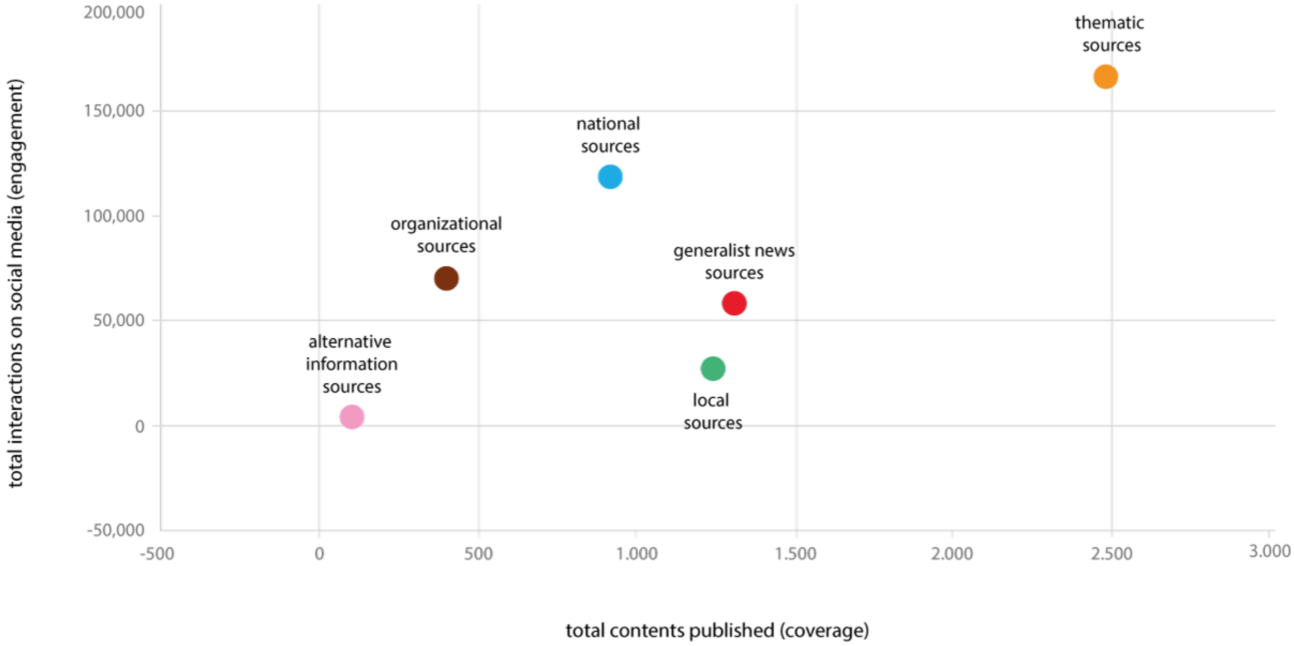
Source name	Category	Contents, N
Ilfattoalimentare.it	Thematic source	256
Ansa.it	National source	250
Meteoweb.eu	Thematic source	194
It.blastingnews.com	Generalist news source	170
Affaritaliani.it	Generalist news source	125
Greenme.it	Thematic source	108
Repubblica.it	National source	107
Greenstyle.it	Thematic source	106
FB Il Fatto Alimentare	Thematic source	90
Lastampa.it	National source	87
Freshplaza.it	Thematic source	81
TW Studio ABR	Thematic source	73
Adnkronos.com	National source	67
Tio.ch	Local source	67
Mainfatti.it	Generalist news source	66

Table 3. Classified source types based on amount of content.

Source type (classified)	Contents, n (%) of the total corpus (N=12,163)
Thematic sources	2480 (20.39)
Generalist news sources	1304 (10.72)
Local sources	1240 (10.19)
National sources	916 (7.53)
Organizational sources	405 (3.33)
Alternative information sources	103 (0.85)
Other sources	26 (0.21)
Total	6474 (53.23)

Regarding the engagement on social media of the content published by the classified sources (Figure 2), the thematic sources achieved the highest number of interactions from online readers, whereas the number of interactions was limited for generalist news and local sources, despite the amount of published content from these sources. National sources ranked in second place, although they published less content related to food risks and safety.

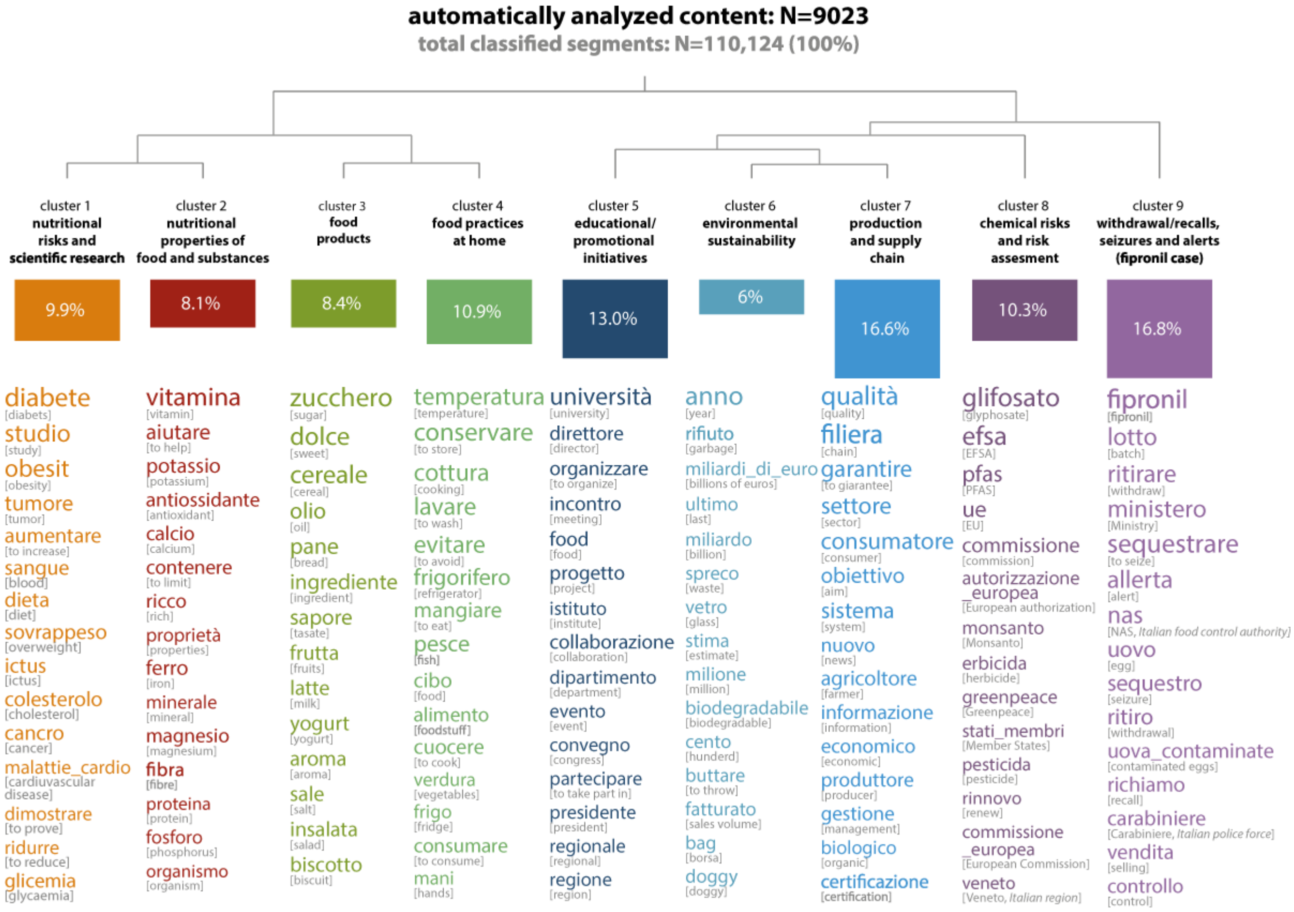
Figure 2. Content and engagement on social media for the classified sources.



Food Risks and Safety Topics That Received Major Coverage and Engagement on Social Media (RQ2)

From the considered subcorpus (n=9023, 74.18% of the total corpus), the automated content analysis turned out nine clusters or lexical worlds (110,124 classified segments, representing 87.08% from the total of 126,463 segments; Figure 3). The hierarchical descending analysis divided these clusters into two subgroups of similar clusters based on content: the former (clusters 1-4) referred to nutritional and biological risks related to the consumption, manipulation, and conservation of food at home, whereas the latter (clusters 5-9) concerned chemical risks and more general or overarching food safety issues, with a strong focus on policies, regulations, and the activities of figures active in this field, such as authorities, companies, institutions, and associations. Multimedia Appendix 6 provides examples of the text segments for each cluster.

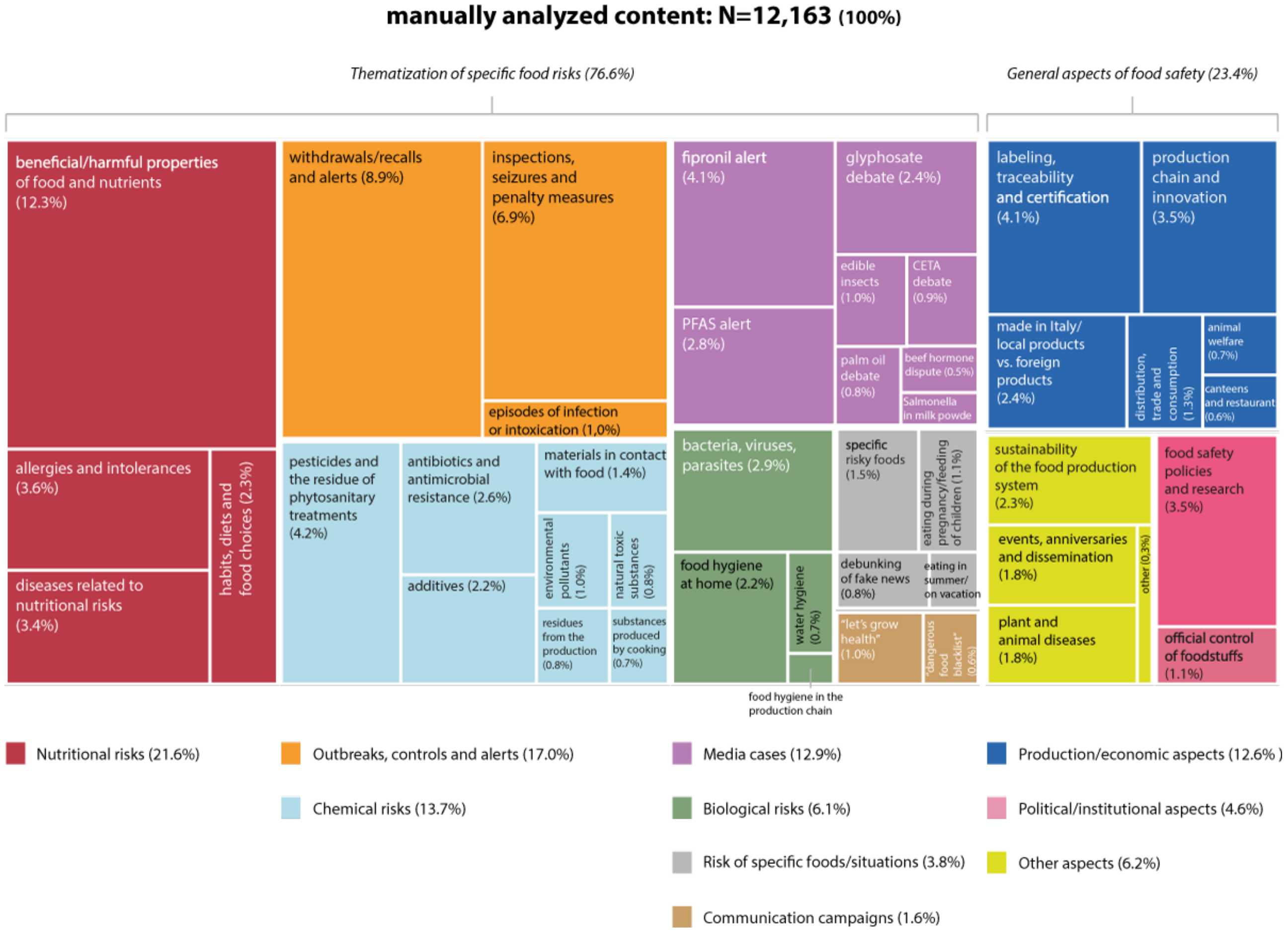
Figure 3. Topics and lexical words.



The manual procedure returned three levels of thematization of food risks in online sources: 45 in-depth categories (second level of tagging), 10 macrocategories (third level of tagging), and (2 macroareas (fourth level of tagging)

The three levels of tagging were matched with the descriptive statistics provided by Web-Live using Qlik Sense to associate the amount of coverage and engagement on social media values with the identified levels of tagging, and to establish which topics emerged from the web monitoring representing food risk and safety issues in the Italian digital ecosystem (Figure 4).

Figure 4. Content for each category identified by the manual content analysis.



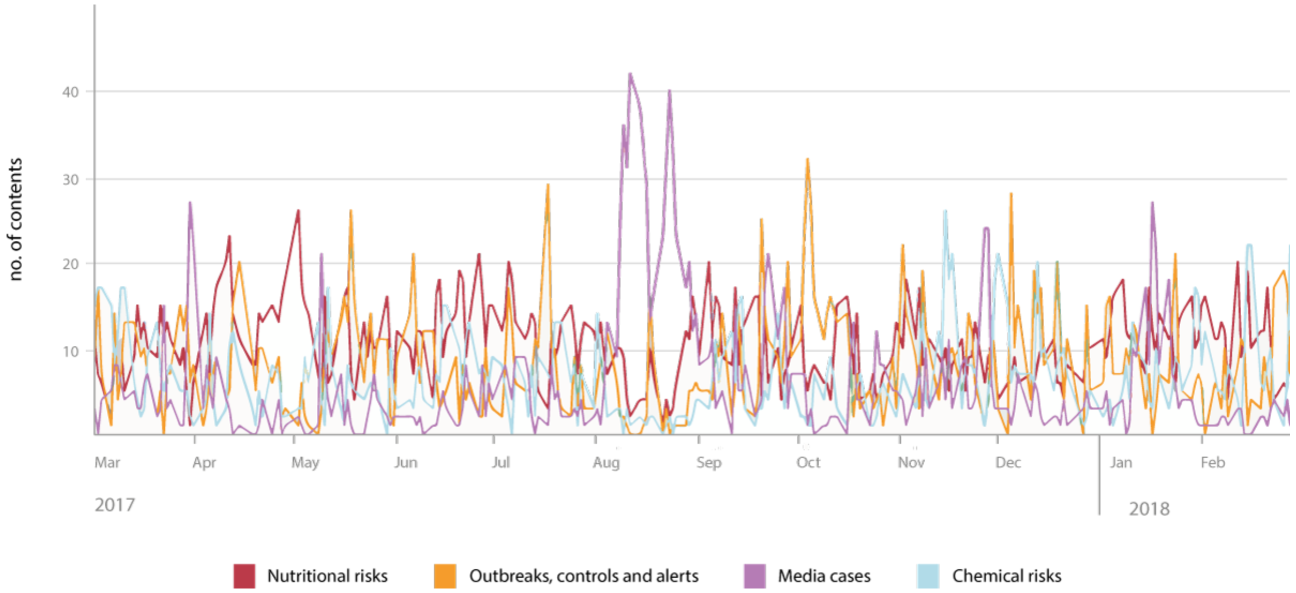
The online coverage of food risks turned out to be composed of two main areas (fourth level of tagging): "thematization of specific food risks" and "general aspects of food safety," which accounted for 76.61% (9318/12,163) and 23.39% (2845/12,163) of the corpus, respectively.

The macroarea "thematization of specific food risks" includes content referring to one or more food risks that stood out well to identified topics. The macroarea "general aspects on food safety" included content that generally refers to food safety as a public health problem, without mentioning or focusing on specific risks. This macroarea was dominated by issues related to the origin and traceability of food products, and the role of certifications and labeling as a means for the consumer to control their own safety of food. In much of this content, products coming from abroad are considered to be unsafe, whereas products of national and local origin are considered safe. This concept is mainly supported by the communications of Italian food producers and trade associations in the agricultural and livestock sector, but is also reinforced by news media and other figures such as politicians and consumer associations. Much of the content is also related to technological developments and innovations in the food production chain, as well as to regulatory policies passed by health authorities and public initiatives created by institutions and associations to ensure food safety.

The macrocategory "nutritional risks" (third level of tagging), with the highest amount of content, emerged as the most important topic associated with food risks, followed by "outbreaks, controls, and alerts." A good level of coverage was also reached by content referring to "chemical risks" and "media cases." The macrocategory "nutritional risks" mainly reports news about "beneficial/harmful properties of food and nutrients" (1497/2623, 57.07% of the content of the macrocategory). This content mentions project results from universities and research centers, focusing on the properties of foods, nutrients, diets, or specific eating habits. This content originates within the scientific community, and online sources translate it into practical advice for the consumer, sometimes reviewing the positive/negative properties of specific foods. Often, this content questions previous studies, highlighting the contradictions in experts' opinions and the partiality of scientific knowledge. The macrocategory "outbreaks, controls, and alerts" reports content related mainly to "withdrawals/recalls and alerts" (1079/2064, 52.28% of the content of the macrocategory) and "inspections, seizures, and penalty measures" (862/2064, 41.76% of the content of the macrocategory). This content derives from the coverage of food alert notifications from the Rapid Alert System for Food and Feed (RASFF) of the European Union, as well as from news that follows the activities and measures adopted by the local health authorities that carry out official controls on the food production and distribution chain. The macrocategory "chemical risks" mainly reports content referring to "pesticides and the residue of phytosanitary treatments" (514/1671, 30.76% of the content of the macrocategory) and "antibiotics and antimicrobial resistance" (319/1671, 19.09% of the content of the macrocategory). Food risks are often associated with intensive farming methods and industrial production, showing a certain distrust of industry and technology. Conversely, the concept of safe and healthy food is associated with the concepts of naturalness and organic farming. The macrocategory “media cases” deals mainly with food alerts referring to the “fipronil alert” (493/1566, 31.48% of the content of the macrocategory), “PFAS (per- and polyfluoroalkyl substances) alert” (345/1566, 22.03% of the content of the macrocategory), and “glyphosate debate” (297/1566, 18.97% of the content of the macrocategory). These cases, and their developments and implications, are generally depicted by adopting the frame of a scandal. This content also importantly covers the public debate on risk management and regulatory policies; activities by risk assessors and managers such as European institutions, national or local health authorities, and private companies are discussed and juxtaposed, each of them feeding the debate with their own views on and perspectives of food risks and safety measures.

Regarding the time distribution of the most commonly covered macrocategories (Figure 5), no significant differences in coverage were found. That is, the attention given by the monitored sources to such topics was rather stable across time, and these topics equally contributed to shaping the source agenda in the reference period. The only sharp increase in coverage was registered during August 2017, which corresponds to the fipronil alert that catalyzed the increased attention of the monitored sources [49], corresponding to a decrease in the coverage of the other major topics.

Figure 5. Trend of the content related to the major food risk topics.



Regarding engagement on social media (Figure 6), the greatest number of interactions (n=398,532) was recorded in November 2017 by a unique content item of the in-depth category “animal welfare” related to a petition promoted through the online portal change.org to support the closure of intensive farming. The other content items that gained a high number of interactions, albeit with significantly lower numbers, were as follows (Figure 7): (1) news content on foods considered to prevent heart attacks (48,704 interactions), published in March 2017, and belonging to the in-depth category “diseases related to nutritional risks”; (2) two distinct news articles published in May 2017 by the environmental association Greenpeace to promote their petition to stop the use of PFAS in the Veneto region (28,182 and 30,561 interactions, respectively) and belonging to the in-depth category “PFAS alert”; and (3) a news article talking about the therapeutic properties of persimmons (56,130 interactions) that was published in October 2017, and belonged to the in-depth category “beneficial/harmful properties of food and nutrients.”

Figure 6. Social media interactions for each category identified by the manual content analysis.

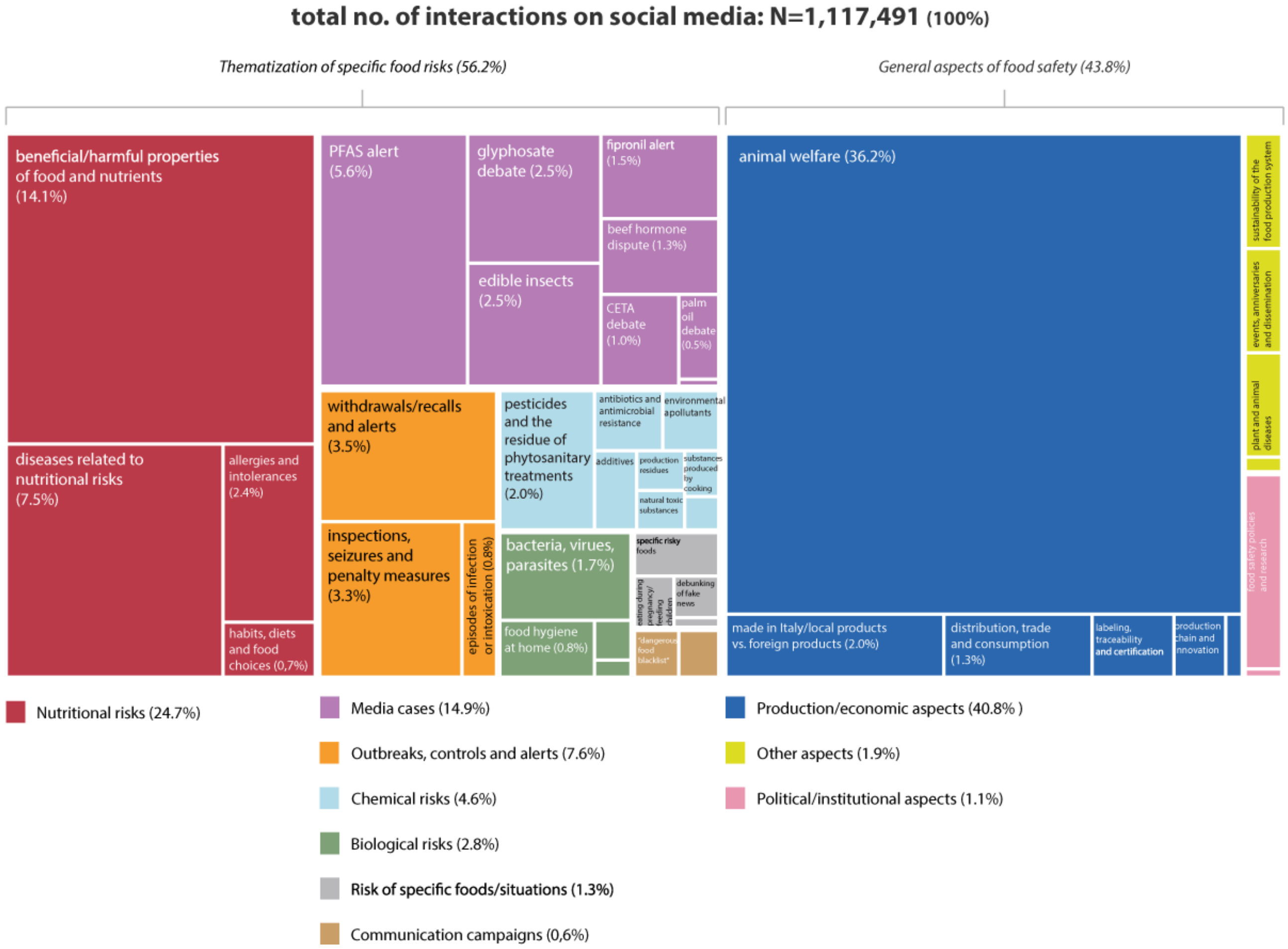
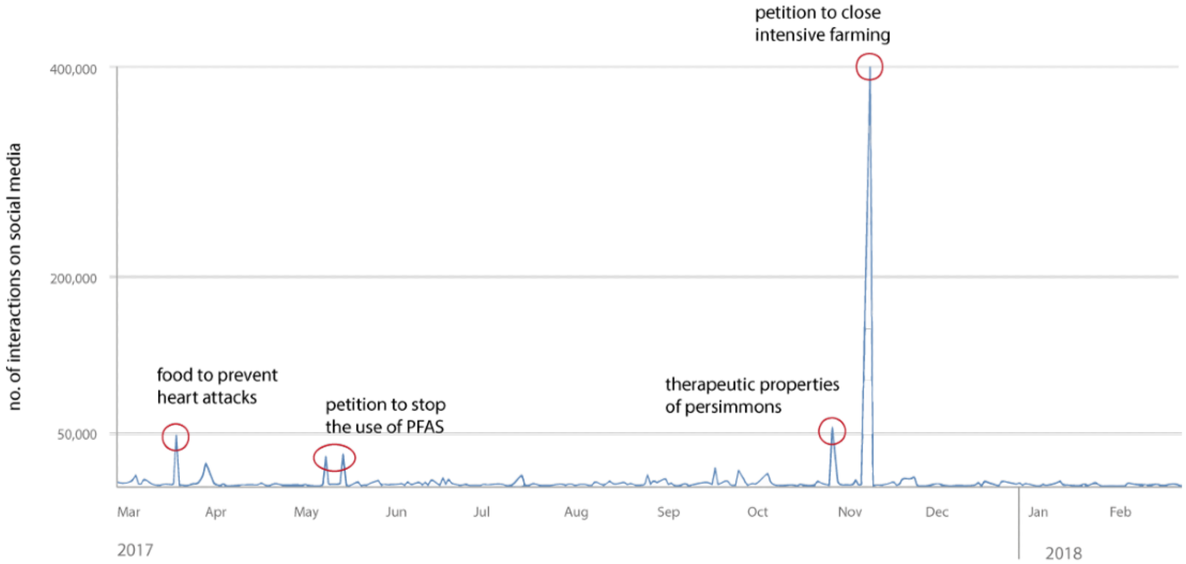


Figure 7. Trend of engagement on social media and major peaks.



Distribution of the Coverage of Food Risk and Safety Topics Among the Monitored Sources (RQ3)

Matching the major topics of the third level of tagging with the amount of content published by the classified sources, we obtained the distribution of these topics by source and how they contributed to composing the source agenda (Table 4).

Thematic sources generally published content about "nutritional risks," whereas generalist news sources and local sources mainly published news related to "outbreaks, controls, and alerts." National sources and organizational sources provided more space to the coverage of "media cases." Alternative information sources talked mostly about "chemical risks."

Table 4. Most widely covered macrocategories and topics (third level of tagging) for each source category.

Source type	Contents, n (%)
Thematic source (N=2408)	
Nutritional risks	825 (33.26)
Chemical risks	364 (15.12)
Outbreaks, controls and alerts	301 (12.50)
Production/economic aspects	259 (10.76)
Media cases	196 (8.14)
Biological risks	177 (7.35)
Specific foods/situations	129 (5.36)
Political/institutional aspects	104 (4.32)
Other aspects	102 (4.32)
Communication campaigns	23 (0.96)
Generalist news sources (N=1304)	
Outbreaks, controls and alerts	331 (25.38)
Nutritional risks	281 (21.55)
Media cases	188 (14.42)
Chemical risks	144 (11.05)
Production/economic aspects	121 (9.28)
Biological risks	77 (5.90)
Other aspects	61 (4.68)
Specific foods/situations	51 (3.91)
Political/institutional aspects	29 (2.22)
Communication campaigns	21 (1.61)
Local sources (N=1240)	
Outbreaks, controls and alerts	315 (25.40)
Media cases	274 (22.10)
Production/economic aspects	160 (12.90)
Nutritional risks	130 (10.48)
Chemical risks	108 (8.71)
Other aspects	92 (7.42)
Political/institutional aspects	58 (4.68)
Biological risks	49 (3.95)
Specific foods/situations	43 (3.47)
Communication campaigns	11 (0.88)
National sources (N=916)	
Media cases	184 (20.1)
Nutritional aspects	167 (18.2)
Chemical risks	150 (16.4)
Production/economic aspects	135 (14.7)
Outbreaks, controls and alerts	85 (9.3)
Other aspects	82 (9.0)
Biological risks	37 (4.0)
Political/institutional aspects	35 (3.8)

Source type	Contents, n (%)
Specific foods/situations	32 (3.5)
Communication campaigns	9 (1.0)
Organizational sources (N=405)	
Media cases	91 (22.5)
Chemical risks	79 (19.5)
Nutritional aspects	73 (18.0)
Production/economic aspects	60 (14.8)
Political/institutional aspects	31 (7.7)
Other aspects	23 (5.7)
Outbreaks, controls and alerts	20 (4.9)
Biological risks	15 (3.7)
Specific foods/situations	10 (2.5)
Communication campaigns	3 (0.7)
Alternative information sources (N=103)	
Chemical risks	36 (35.0)
Nutritional risks	33 (32.0)
Media cases	13 (12.6)
Outbreaks, controls and alerts	9 (8.7)
Biological risks	5 (4.9)
Other aspects	3 (2.9)
Production/economic aspects	3 (2.9)
Specific foods/situations	1 (1.0)

Discussion

Principal Findings

The results of this study show that food risk and safety issues are widely covered by the monitored online sources. During the period under study, the sources continued publishing content and their interest in the theme was stable over time, apart from media peaks corresponding to major food scares. This finding resonates with previous results referring to the coverage of food risks by daily newspapers in Italy [13,40]. The results also show that the online sources do not talk about food risks and safety only during a crisis but, importantly, also during periods of peace, when a great part of the coverage was also tracked. This pattern enables online readers to realize the complexity of this field in terms of both the topics and the actors concerned, even when there is no imminent or direct risk for people.

RQ1: Which Online Information Sources Have Covered Food Risks the Most and Received Major Engagement on Social Media?

Notwithstanding the growing role of social media as a source of information [6], a great amount of the content on food risk and safety issues is still published on websites and news portals, which differ in terms of source type and editorial line. Owing to the monitoring and analysis of content from a large set of sources, our data reveal that traditional sources (eg, national and local news media) have been joined by other types of sources that give voice to a variety of authors, such as thematic and generalist news sources. Indeed, food risk and safety issues are also covered by numerous sources that have published one or more items related to the topic in an occasional way and without making it a distinctive sign of their editorial line. This highlights how actors not primarily involved in the management and communication of food risks can actually play the role of alternative online science communicators. This result resonates with the increasingly consolidated phenomenon that sees activists and advocates in the health and environmental fields competing with other societal actors for public attention to issues [50], and scientists are not always seen by the relevant audience as figures who can be easily related to, especially in social media communication [51]. Our data about engagement on social media peaks also confirm this trend.

The thematic sources ranked first for the amount of published content on food risks and reached the highest level of engagement on social media. This is probably because these sources offer more specialized content on the subject, targeting the attention of readers with a specific interest in this content who are therefore more likely to share and discuss it with their social network. Importantly, local and national sources (mainly news media websites) still cover food risk and safety topics: authors belonging to national sources are among the most active in terms of published content, and their informative role is still greater in cases of food emergencies (“media cases” tag). Interestingly, the second type of source in terms of the amount of published content is online sources of generalist news. This category includes national news portals, news aggregator websites, citizen journalism platforms, and related social accounts, representing the change that news production practices have undergone in the last few decades [52]. A smaller amount of content was published by sources of alternative information that generally publish fake news or content of uncertain reliability. However, this content obtained almost no engagement on social media, thereby limiting the potential dissemination of inaccurate news.

RQ2: Which Food Risk Topics Have Received Major Attention in Terms of Coverage and Engagement on Social Media?

The “food risk” theme consists of several smaller topics that can be divided into two major categories: “food risks in everyday life” and “food risk governance.” The former category discusses nutritional and biological risks that people can manage on their own and upon which they can have direct control by means of simple actions such as diet choices and food preparation practices. The latter category refers to national and international activities to practice food safety and to ensure that the debate about risk management and related policies is made accessible to the public. This latter category includes chemical risks that are generally perceived as not directly controllable in everyday life, as the responsibility for such risks is considered to be external and technology-driven.

The prevalence of content about nutrition resonates with the fundamental role that nutritional science has assumed in shaping food meanings and practices, and nutrition experts’ presence and advice are now prolific in the media [53]. The results show that this also occurs in the Italian online communication landscape, where nutritional information is widely covered by the monitored sources and has also gained considerable interest from readers, with the macrocategory “nutritional risks” being the topic with the second-highest value of engagement on social media. Indeed, this topic is very close to consumers’ everyday lives and possibly fulfills their information needs. Notwithstanding, the way this topic is discussed (ie, either corroborating or questioning previous research in this field) prevents readers from easily interpreting the information and, consequently, appropriately managing the risk itself in everyday life because they receive contradictory messages over time.

The significant number of mentions referring to the macrocategory “outbreaks, controls, and alerts” proves that news on withdrawals, recalls, inspections discovering a lack of safety measures, and foodborne outbreaks is regularly published online but normally does not receive particular amplification in terms of coverage and readers’ interactions. This content enters the national sources agenda when factors such as a high number of people concerned and the presence of victims or the unknown/new nature of the risk intervene to increase the newsworthiness of the event. This mechanism actually characterizes the macrocategory “media cases” and its related in-depth categories such as the “fipronil alert” [49]. This finding reinforces the role of external factors in influencing the newsworthiness and the coverage of risks, notwithstanding the characteristics of the hazards themselves.

The widespread interest of the monitored sources in “chemical risks” matches with what consumers report as their major concerns about food [3,54], which has also been explored in the Italian context [2]. Actually, the “media cases” that received major coverage had chemical origins and referred to pesticides (fipronil and glyphosate) and environmental pollutants (PFAS).

Despite the diffusion of foodborne diseases at home [55] and the need to help consumers prevent them [56-58], the results show that “biological risks” receives minor coverage compared to nutritional and chemical risks. Differences in the nature of these risks justify this coverage pattern. Biological risks generally refer to a consolidated knowledge of pathogens and related foodborne diseases, with well-defined instructions to prevent them. In contrast, chemical risks cannot be so easily perceived and managed, and their controversial nature makes them more attractive to cover. Indeed, the way these risks are reported gives rise to the many, often contradictory, opinions supported by the diverse figures involved (eg, risk assessors, risk managers, health authorities, consumer associations). Biological risks offer much fewer possibilities for using contradictory opinions to report facts, at least in the reference period of this study, in which no significant food crisis with biological origins occurred.

RQ3: Is There Any Difference in the Coverage of Food Risk and Safety Topics Among the Monitored Sources?

The results further highlight some variability among the agendas of the monitored sources. National and organizational sources generally cover food risk topics when they are “media cases”: the news-making criteria for these sources is thus the occurrence of “risk events” [14], where the story can be featured in terms of scandal, emergency, and victims, and generally goes beyond local boundaries. Despite the lower amount of this content, it reached a good level of engagement on social media. This means that people are more likely to interact with content referring to major national events that possibly refer to a threat to public health. Generalist news sources and local sources give more space to issues related to “outbreaks, controls, and alerts”: these issues generally originate as local facts that occur in well-defined territories, thus justifying the widespread interest of local sources in this type of content.

Previous studies focusing on the coverage of food risks by online daily newspapers showed that the great majority of RASFF data did not receive specific mention, apart from food alerts [40]. Notably, this study not only confirms the preference of national sources for “food risk events” but also shows that a great number of alerts mentioned in the RASFF or more general content referring to official controls on food by health authorities are actually mentioned by online thematic, generalist, and local sources through which people can come to know about such information.

Limitations

A first important limitation of this study refers to the inherent nature of big data that influences digital social research in terms of sample representativeness and access [35]. Our data were retrieved by a commercial web-monitoring app, whose algorithms are not completely disclosable for reasons of commercial protection [22]. The sampling method also suffered from access restriction from social media platforms, which give only partial access to their data [35,59]. To overcome these limitations and solve any opacity, we included the monitoring profile used to retrieve the analyzed content (Multimedia Appendix 1). Notably, the tracking of content related to the fipronil alert proved the sensitivity of the monitoring profile and the selection procedure, as the keyword "fipronil" was not included in the monitoring profile. In addition, the results need to be interpreted and contextualized within the limit of the 50 validated daily mentions: a different limit might have been liable to generate different coverage patterns.

The analysis of the web-monitoring data texts was performed with both automated and manual content analysis to obtain a deeper interpretation. We believe that both procedures are needed to better translate texts into sound and actionable information. The automated analysis enabled us to give robustness to the manual procedure: it provided us with the main branches referring to food risks and safety that worked as a starting point to guide the manual procedure. In turn, although it is a highly time-consuming activity, the manual analysis yielded detailed insights, and allowed us to understand the granularity and faceted nature of food risk and safety communication.

Conclusions and Further Research

The combination of web monitoring, content analysis, and data visualization techniques proposed in this study was proven to be a viable approach to understand when the media pick up on an issue that does not actually jeopardize public health but generates a great deal of coverage, as well as where and from whom there is silence during an important issue or outbreak. The presence or lack of coverage of specific risks and topics in the online communication of food risks, as an output of web-monitoring data, can help communication practitioners in health agencies better hone their communication strategies and interventions.

The monitoring of a broad spectrum of online sources allowed for the tracking and interpretation of the interplay between science and society for a wider understanding of the mechanisms underlying food risk communication. This study demonstrates how the online information ecosystem of risk communication has changed and assisted with a proliferation of information sources that work as new mediators [60], the reliability and standing of which might sometimes be questionable. It is very clear that there is room for expert figures (eg, governmental institutions, food safety agencies, health authorities, research centers) to deliver food risk information using the online environment. Finally, communication practitioners could work in a closer relationship with the editorial offices of thematic sources, which talked about food risks more than the other types of sources. In this way, the scientific community could have greater visibility and coverage in the online environment, guaranteeing scientifically sound information.

Acknowledgments

This work was funded by the Italian Ministry of Health (grant number RC IZSVe 04/2014).

The authors would like to thank Riccardo Di Marcantonio and Maria Grazia Graziani, and their Extreme team for support in gathering and analyzing the data. The authors also thank Carmen Losasso for her contribution to the development of the monitoring profile.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Rules of keywords and logical operators used to configure the web monitoring app Web-Live [22] to monitor food risks and food safety topics in the Italian online media.
[TXT File , 20 KB - jmir_v22i12e23438_app1.txt]

Multimedia Appendix 2
Description of the web monitoring app Web-Live.
[DOCX File , 16 KB - jmir_v22i12e23438_app2.docx]

Multimedia Appendix 3
Definition of the topic categories identified with the manual content analysis.
[DOCX File , 23 KB - jmir_v22i12e23438_app3.docx]

Multimedia Appendix 4
Categories of online information sources.

[DOCX File , 13 KB - jmir_v22i12e23438_app4.docx]

Multimedia Appendix 5
Texts and statistics related to the final validated and analyzed sample.
[XLSX File (Microsoft Excel File), 22790 KB - jmir_v22i12e23438_app5.xlsx]

Multimedia Appendix 6
Examples of text segments for each cluster in the food risk corpus.
[DOCX File , 16 KB - jmir_v22i12e23438_app6.docx]

References

1. Röhr A, Lüddecke K, Drusch S, Müller M, Alvensleben R. Food quality and safety—consumer perception and public health concern. Food Control 2005 Oct;16(8):649-655. [doi: 10.1016/j.foodcont.2004.06.001]
2. Tiozzo B, Mari S, Ruzza M, Crovato S, Ravarotto L. Consumers' perceptions of food risks: A snapshot of the Italian Triveneto area. Appetite 2017 Apr 01;111:105-115. [doi: 10.1016/j.appet.2016.12.028] [Medline: 28024884]
3. Food safety in the EU: Special Eurobarometer Wave EB91. European Food Safety Authority. 2019 Apr. URL: https://www.efsa.europa.eu/sites/default/files/corporate_publications/files/Eurobarometer2019_Food-safety-in-the-EU_Full-report.pdf [accessed 2020-11-20]
4. Panagiotopoulos P, Barnett J, Bigdeli AZ, Sams S. Social media in emergency management: Twitter as a tool for communicating risks to the public. Technol Forecast Soc Change 2016 Oct;111:86-96. [doi: 10.1016/j.techfore.2016.06.010]
5. Tiozzo B, Pinto A, Mascarello G, Mantovani C, Ravarotto L. Which food safety information sources do Italian consumers prefer? Suggestions for the development of effective food risk communication. J Risk Res 2018 Feb 19;22(8):1062-1077. [doi: 10.1080/13669877.2018.1440414]
6. Kuttschreuter M, Rutsaert P, Hilverda F, Regan Á, Barnett J, Verbeke W. Seeking information about food-related risks: The contribution of social media. Food Qual Pref 2014 Oct;37:10-18. [doi: 10.1016/j.foodqual.2014.04.006]
7. Zhang Y, Sun Y, Kim Y. The influence of individual differences on consumer's selection of online sources for health information. Comput Hum Behav 2017 Feb;67:303-312. [doi: 10.1016/j.chb.2016.11.008]
8. Hilverda F, Kuttschreuter M, Giebels E. Social media mediated interaction with peers, experts and anonymous authors: Conversation partner and message framing effects on risk perception and sense-making of organic food. Food Qual Pref 2017 Mar;56:107-118. [doi: 10.1016/j.foodqual.2016.09.003]
9. Rutsaert P, Regan A, Pieniak Z, McConnon A, Moss A, Wall P, et al. The use of social media in food risk and benefit communication. Trends Food Sci Technol 2013 Mar;30(1):84-91. [doi: 10.1016/j.tifs.2012.10.006]
10. Brossard D, Scheufele DA. Social science. Science, new media, and the public. Science 2013 Jan 04;339(6115):40-41. [doi: 10.1126/science.1232329] [Medline: 23288529]
11. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Bus Horizon 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
12. Huurne ET, Gutteling J. Information needs and risk perception as predictors of risk information seeking. J Risk Res 2008 Oct;11(7):847-862. [doi: 10.1080/13669870701875750]
13. Carducci A, Alfani S, Sassi M, Cinini A, Calamusa A. Mass media health information: quantitative and qualitative analysis of daily press coverage and its relation with public perceptions. Patient Educ Couns 2011 Mar;82(3):475-478. [doi: 10.1016/j.pec.2010.12.025] [Medline: 21288683]
14. Driedger SM, Jardine CG, Boyd AD, Mistry B. Do the first 10 days equal a year? Comparing two Canadian public health risk events using the national media. Health Risk Soc 2009 Feb;11(1):39-53. [doi: 10.1080/13698570802537011]
15. Strekalova YA. Emergent health risks and audience information engagement on social media. Am J Infect Control 2016 Mar 01;44(3):363-365. [doi: 10.1016/j.ajic.2015.09.024] [Medline: 26559738]
16. Chang Y, Ku C, Chen C. Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. Int J Inf Manage 2019 Oct;48:263-279. [doi: 10.1016/j.ijinfomgt.2017.11.001]
17. Galli B. Web listening: conoscere per agire. Milano: Franco Angeli; 2018.
18. Jeong B, Yoon J, Lee J. Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. Int J Inf Manage 2019 Oct;48:280-290. [doi: 10.1016/j.ijinfomgt.2017.09.009]
19. Shirdastian H, Laroche M, Richard M. Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter. Int J Inf Manage 2019 Oct;48:291-307. [doi: 10.1016/j.ijinfomgt.2017.09.007]
20. Thomaz GM, Biz AA, Bettoni EM, Mendes-Filho L, Buhalis D. Content mining framework in social media: A FIFA world cup 2014 case analysis. Inf Manag 2017 Sep;54(6):786-801. [doi: 10.1016/j.im.2016.11.005]
21. Corley CD, Cook DJ, Mikler AR, Singh KP. Text and structural data mining of influenza mentions in Web and social media. Int J Environ Res Public Health 2010 Feb;7(2):596-615 [FREE Full text] [doi: 10.3390/ijerph7020596] [Medline: 20616993]

22. Lipizzi C, Iandoli L, Ramirez Marquez JE. Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter streams. Int J Inf Manage 2015 Aug;35(4):490-503. [doi: 10.1016/j.ijinfomgt.2015.04.001]
23. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. Int J Inf Manage 2015 Apr;35(2):137-144. [doi: 10.1016/j.ijinfomgt.2014.10.007]
24. Guo L, Vargo CJ, Pan Z, Ding W, Ishwar P. Big Social Data Analytics in Journalism and Mass Communication. Journalism Mass Commun Quart 2016 Apr 12;93(2):332-359. [doi: 10.1177/1077699016639231]
25. van der Meer TG. Automated content analysis and crisis communication research. Public Relat Rev 2016 Dec;42(5):952-961. [doi: 10.1016/j.pubrev.2016.09.001]
26. Junqué de Fortuny E, De Smedt T, Martens D, Daelemans W. Media coverage in times of political crisis: A text mining approach. Expert Syst Apps 2012 Oct;39(14):11616-11622. [doi: 10.1016/j.eswa.2012.04.013]
27. Karlsson M, Sjøvaag H. Content Analysis and Online News. Dig Journalism 2015 Oct 28;4(1):177-192. [doi: 10.1080/21670811.2015.1096619]
28. Mazzocut M, Truccolo I, Antonini M, Rinaldi F, Omero P, Ferrarin E, et al. Web Conversations About Complementary and Alternative Medicines and Cancer: Content and Sentiment Analysis. J Med Internet Res 2016 Jun 16;18(6):e120 [FREE Full text] [doi: 10.2196/jmir.5521] [Medline: 27311444]
29. Bekkers V, Edwards A, de Kool D. Social media monitoring: Responsive governance in the shadow of surveillance? Gov Inf Quart 2013 Oct;30(4):335-342. [doi: 10.1016/j.giq.2013.05.024]
30. He W, Zha S, Li L. Social media competitive analysis and text mining: A case study in the pizza industry. Int J Inf Manage 2013 Jun;33(3):464-472. [doi: 10.1016/j.ijinfomgt.2013.01.001]
31. Yeung RM, Morris J. Food safety risk. Br Food J 2001 Apr;103(3):170-187. [doi: 10.1108/00070700110386728]
32. Mol APJ, Bulkeley H. Food risks and the environment: changing perspectives in a changing social order. J Environ Policy Plan 2002 Sep;4(3):185-195. [doi: 10.1002/jepp.109]
33. Dreyer M, Renn O. Risk Analysis: Risk Communication: Chemical Hazards. In: Motarjemi Y, editor. Encyclopedia of Food Safety. Cambridge: Academic Press; 2014:122-126.
34. Trilling D, Jonkman JGF. Scaling up Content Analysis. Commun Method Meas 2018 Mar 14;12(2-3):158-174. [doi: 10.1080/19312458.2018.1447655]
35. Veltri G. Digital Social Research. Cambridge: Polity; 2019.
36. Sormanen N, Rohila J, Lauk E, Uskali T, Jouhki J, Penttinen M. Chances and Challenges of Computational Data Gathering and Analysis. Dig Journalism 2015 Oct 29;4(1):55-74. [doi: 10.1080/21670811.2015.1096614]
37. Tao D, Yang P, Feng H. Utilization of text mining as a big data analysis tool for food science and nutrition. Comp Rev Food Sci Food Safe 2020 Feb 16;19(2):875-894. [doi: 10.1111/1541-4337.12540]
38. Fusté-Forné F, Masip P. Food in journalistic narratives: A methodological design for the study of food-based contents in daily newspapers. Int J Gastron Food Sci 2018 Dec;14:14-19. [doi: 10.1016/j.ijgfs.2018.08.004]
39. Qiang L, Wen L, Jing W, Yue D. Application of content analysis in food safety reports on the Internet in China. Food Control 2011 Feb;22(2):252-256. [doi: 10.1016/j.foodcont.2010.07.005]
40. Tiozzo B, Pinto A, Neresini F, Sbalchiero S, Parise N, Ruzza M, et al. Food risk communication: analysis of the media coverage of food risk on Italian online daily newspapers. Qual Quant 2019 May 30;53(6):2843-2866. [doi: 10.1007/s11135-019-00897-3]
41. Günther E, Quandt T. Word Counts and Topic Models. Dig Journalism 2015 Oct 13;4(1):75-88. [doi: 10.1080/21670811.2015.1093270]
42. Bolasco S. TaLTaC 2. Sviluppi, esperienze ed elementi essenziali di analisi automatica dei testi. Milano: LED Edizioni Universitarie; 2010.
43. Ratinaud P. Iramuteq. 2009. URL: http://www.iramuteq.org/front-page/presentation_view [accessed 2020-11-20]
44. Max R. Les "mondes lexicaux" et leur 'logique' à travers l'analyse statistique d'un corpus de récits de cauchemars. lsoc 1993;66(1):5-39. [doi: 10.3406/lsoc.1993.2632]
45. Neuendorf K. The content analysis guidebook. London: Sage; 2016.
46. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
47. Stieglitz S, Mirbabaie M, Ross B, Neuberger C. Social media analytics – Challenges in topic discovery, data collection, and data preparation. Int J Inf Manage 2018 Apr;39:156-168. [doi: 10.1016/j.ijinfomgt.2017.12.002]
48. Tiozzo B, Ruzza M, Rizzoli V, D'Este L, Giaretta M, Ravarotto L. Food risks by Italian online information sources. 2020. URL: https://sense.izsvenezie.it/pub/extensions/food-risks-web-2017-2018/food-risks-web-2017-2018.html [accessed 2020-11-20]
49. Ruzza M, Tiozzo B, Rizzoli V, Giaretta M, D'Este L, Ravarotto L. Food Risks on the Web: Analysis of the 2017 Fipronil Alert in the Italian Online Information Sources. Risk Anal 2020 Oct;40(10):2071-2092. [doi: 10.1111/risa.13533] [Medline: 32557616]
50. Trench B. Towards an Analytical Framework of Science Communication Models. In: Communicating Science in Social Contexts. Dordrecht: Springer; 2008:119-135.

51. Bucchi M. The "Greta Thunberg Effect": some reflections on communication. 2019. URL: https://www.esof.eu/en/news-reader/the-greta-thunberg-effect-some-reflections-on-communication-427.html [accessed 2019-07-22]
52. Widholm A. Tracing Online News in Motion. Dig Journalism 2015 Oct 27;4(1):24-40. [doi: 10.1080/21670811.2015.1096611]
53. Dodds A, Chamberlain K. The problematic messages of nutritional discourse: A case-based critical media analysis. Appetite 2017 Jan 01;108:42-50. [doi: 10.1016/j.appet.2016.09.021] [Medline: 27659870]
54. Bearth A, Cousin M, Siegrist M. Investigating novice cooks' behaviour change: Avoiding cross-contamination. Food Control 2014 Jun;40:26-31. [doi: 10.1016/j.foodcont.2013.11.021]
55. European Food Safety Authority European Centre for Disease Prevention Control (EFSAECDC). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017. EFSA J 2018 Dec;16(12):e05500 [FREE Full text] [doi: 10.2903/j.efsa.2018.5500] [Medline: 32625785]
56. Evans EW, Redmond EC. Behavioral risk factors associated with listeriosis in the home: a review of consumer food safety studies. J Food Prot 2014 Mar;77(3):510-521. [doi: 10.4315/0362-028X.JFP-13-238] [Medline: 24674447]
57. Geppert J, Schulze Struchtrup S, Stamminger R, Haarhoff C, Ebert V, Koch S, et al. Food safety behavior observed in German TV cooking shows. Food Control 2019 Feb;96:205-211. [doi: 10.1016/j.foodcont.2018.09.017]
58. Ruby GE, Ungku Zainal Abidin UF, Lihan S, Jambari NN, Radu S. Predicting intention on safe food handling among adult consumers: A cross sectional study in Sibu district, Malaysia. Food Control 2019 Dec;106:106696. [doi: 10.1016/j.foodcont.2019.06.022]
59. Shan L, Regan A, De Brún A, Barnett J, van der Sanden MCA, Wall P, et al. Food crisis coverage by social and traditional media: A case study of the 2008 Irish dioxin crisis. Public Underst Sci 2014 Nov;23(8):911-928. [doi: 10.1177/0963662512472315] [Medline: 23825281]
60. Smith A, Parrino L, Vrbos D, Nicolini G, Bucchi M, Carr M, et al. Communicating to and engaging with the public in regulatory science. EFSA J 2019 Jul;17(Suppl 1):e170717 [FREE Full text] [doi: 10.2903/j.efsa.2019.e170717] [Medline: 32626454]

Abbreviations

PFAS: per- and polyfluoroalkyl substances
RASFF: Rapid Alert System for Food and Feed

Edited by R Kukafka; submitted 26.08.20; peer-reviewed by E Andrikopoulou, C Reis; comments to author 20.09.20; revised version received 21.09.20; accepted 19.10.20; published 14.12.20.

Please cite as:
Tiozzo B, Ruzza M, Rizzoli V, D'Este L, Giaretta M, Ravarotto L
Biological, Chemical, and Nutritional Food Risks and Food Safety Issues From Italian Online Information Sources: Web Monitoring, Content Analysis, and Data Visualization
J Med Internet Res 2020;22(12):e23438
URL: https://www.jmir.org/2020/12/e23438
doi:10.2196/23438
PMID:33315018

Original Paper

Web-Based Privacy-Preserving Multicenter Medical Data Analysis Tools Via Threshold Homomorphic Encryption: Design and Development Study

Yao Lu[1*], BSc; Tianshu Zhou[1*], PhD; Yu Tian[1], PhD; Shiqiang Zhu[2], PhD; Jingsong Li[1,2], PhD

[1]Engineering Research Center of EMR and Intelligent Expert System, Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

[2]Zhejiang Lab, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Jingsong Li, PhD
Engineering Research Center of EMR and Intelligent Expert System
Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Science
Zhejiang University
38 Zheda Road
Hangzhou, 310027
China
Phone: 86 571 87951564
Email: ljs@zju.edu.cn

Abstract

Background: Data sharing in multicenter medical research can improve the generalizability of research, accelerate progress, enhance collaborations among institutions, and lead to new discoveries from data pooled from multiple sources. Despite these benefits, many medical institutions are unwilling to share their data, as sharing may cause sensitive information to be leaked to researchers, other institutions, and unauthorized users. Great progress has been made in the development of secure machine learning frameworks based on homomorphic encryption in recent years; however, nearly all such frameworks use a single secret key and lack a description of how to securely evaluate the trained model, which makes them impractical for multicenter medical applications.

Objective: The aim of this study is to provide a privacy-preserving machine learning protocol for multiple data providers and researchers (eg, logistic regression). This protocol allows researchers to train models and then evaluate them on medical data from multiple sources while providing privacy protection for both the sensitive data and the learned model.

Methods: We adapted a novel threshold homomorphic encryption scheme to guarantee privacy requirements. We devised new relinearization key generation techniques for greater scalability and multiplicative depth and new model training strategies for simultaneously training multiple models through x-fold cross-validation.

Results: Using a client-server architecture, we evaluated the performance of our protocol. The experimental results demonstrated that, with 10-fold cross-validation, our privacy-preserving logistic regression model training and evaluation over 10 attributes in a data set of 49,152 samples took approximately 7 minutes and 20 minutes, respectively.

Conclusions: We present the first privacy-preserving multiparty logistic regression model training and evaluation protocol based on threshold homomorphic encryption. Our protocol is practical for real-world use and may promote multicenter medical research to some extent.

(J Med Internet Res 2020;22(12):e22555) doi:10.2196/22555

KEYWORDS

machine learning; confidentiality; threshold homomorphic encryption; logistic regression

Introduction

Background

In recent years, researchers have proposed strong requirements for the quality of medical research as it continues to progress, which has promoted the development of multicenter research. Compared with single-center research, multicenter research has many significant advantages, including enabling specific analyses for which no single institution has sufficient data, such as on a rare disease; providing medical data from different locations with diverse demographics, which increases the reproducibility and generalizability of the research; and generating pooled medical data that enables new discoveries that cannot be elucidated from any individual data set [1,2]. In addition, the development of multicenter medical research has accelerated the translation of research outcomes into clinical practice and strengthened collaborations among institutions [2,3].

However, data sharing during multicenter research may increase privacy security risks. As medical data are highly sensitive, the leakage of sensitive information will lead to severe consequences, such as financial loss, social discrimination, and unauthorized data abuse, which can harm both patients and medical institutions [4]. As a result, many medical institutions are unwilling to share their data despite the aforementioned benefits, which hinders the collaborative benefits of multicenter research. To solve this problem, a framework is urgently needed to support multicenter medical research efficiently while preventing the leakage of sensitive information.

Prior Work

Logistic regression is a widely used machine learning approach in various medical applications, such as prognostic prediction, disease diagnosis, and decision-making support [5]. For example, Abdolmaleki et al [6] used logistic regression to predict the outcome of biopsy in breast cancer and obtained 90% accuracy. Many solutions have been developed to address privacy-preserving logistic regression. Some use intermediary statistics to train a model without accessing the raw data; however, these methods remain vulnerable to statistical attack when a particular criterion holds true for only one sample [7-9]. Other researchers use homomorphic encryption to protect privacy during model training, which is similar to that used in this study [10-19]. Homomorphic encryption technology provides rigorous protection for sensitive information and enables the computation of information in an encrypted format and is, therefore, a potential candidate for secure logistic regression model training. However, unlike our solution, these homomorphic encryption–based solutions yield only sets of parameters, and there are no methods to evaluate the trained model in a secure manner. Furthermore, these methods use a single public and secret key, meaning that all the research data may be exposed to anyone who holds the secret key, limiting the application of these solutions in real-life scenarios. In the current literature, the works most similar to ours are those of Emam et al [18] and Jiang et al [19], which attempt to avoid information leak using methods that differ from ours. Emam et al [18] kept the data local to the corresponding data providers and used the Paillier scheme to deal with intermediate values. However, because the public and secret keys are stored at the central unit, when multiple parties collude with the central unit, some meaningful information about the other parties' sensitive data may be revealed to them [18]. Jiang et al [19] proposed a hybrid cryptographic method that uses a software guard extensions (SGX) enclave to securely generate and store the secret key in a trusted cloud. As the cloud server is shared among different users, it is more likely to be attacked. Considering the rapid development of attack methods toward SGX, including a recently proposed method capable of stealing the enclave secret to subvert the confidentiality of SGX, placement of the secret key in the cloud is not secure [20]. Once the attackers break through the SGX's guard, they will be able to obtain the secret key and decrypt all the sensitive information stored on the cloud, leading to a severe outcome.

Multikey homomorphic encryption, first proposed by López-Alt et al [21], allows computations on ciphertexts under different secret keys, which makes the method suitable for secure multicenter research. However, the scheme proposed in the study by Lopez et al [21] is based on the Nth degree truncated polynomial ring units cryptosystem, where if we obtain a result computed from ciphertexts under different keys, we will need to decrypt the result by the product of all involved secret keys, allowing for only a very limited number of parties before the decryption error grows too large to obtain the correct plaintext result. Another multikey homomorphic encryption method, called threshold homomorphic encryption, allows many more parties to participate without resulting in an excessively large decryption error; however, the noise generated in the relinearization is still very large and grows quadratically with the number of parties, which would have a negative effect on the multiplicative depth [22].

Objectives

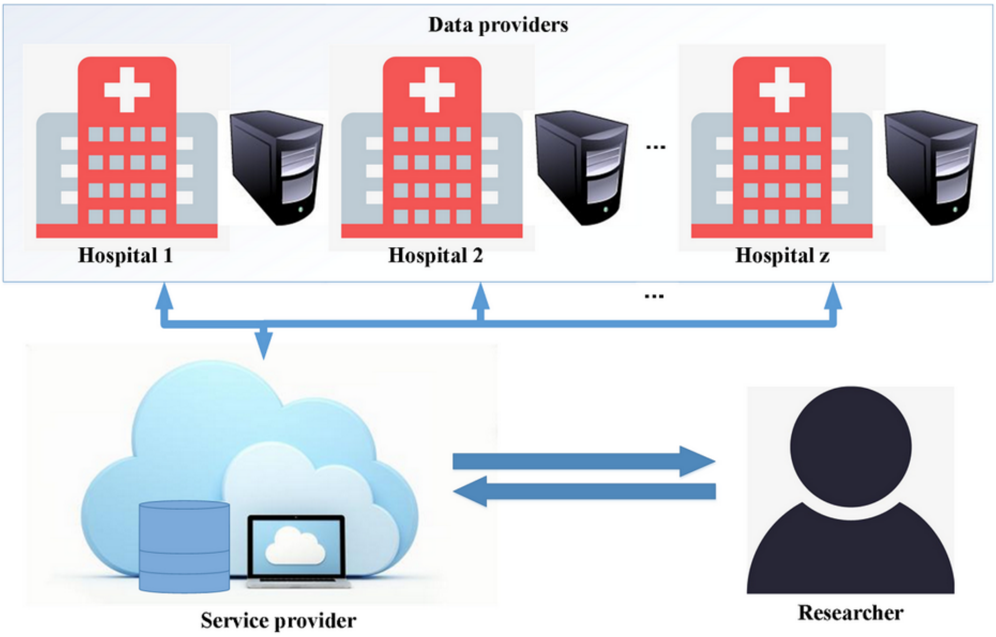
In this study, we propose a privacy-preserving multicenter research protocol using secure logistic regression, consisting of 3 primary entities: researchers, a service provider, and data providers, in which medical data are horizontally distributed. Our proposed protocol supports not only model training but also the evaluation of the trained model in a secure manner. The protocol guarantees the privacy of both the sensitive data for the data providers and the trained model for the researchers during model training and trained model evaluation. To satisfy privacy requirements, we apply threshold homomorphic encryption and propose a new relinearization key generation process that increases scalability and multiplicative depth. The proposed protocol has been implemented and tested with simulated real-life scenarios. The experimental results demonstrate that our protocol is efficient and practical for real-world applications.

Methods

Overview of the Presented Protocol

Our proposed protocol includes 3 primary entities as shown below. The architecture of the proposed protocol is shown in Figure 1.

Figure 1. The architecture of the proposed protocol, containing 3 entities: data providers, a service provider, and researchers.



Data Providers

These include institutions (eg, hospitals) who hold medical data and are willing to provide these data to the service provider for public use so long as the privacy of the data is preserved. To share medical data, the data providers must obtain patient consent if the local law requires so. Upon receiving the researchers' requests from the service provider, the data providers can decide whether to accept or refuse. To allow researchers to obtain correct research data, all data providers must implement data standardization to transform the data into a common format, such as the Observational Medical Outcomes Partnership common data model from the Observational Health Data Sciences and Informatics collaborative [23].

Service Provider

This refers to an entity that (1) provides storage for encrypted data and research information, (2) performs the most computationally expensive part of the privacy-preserving logistic regression, and (3) performs information transfer among the data providers, the service provider, and the researchers. In addition, an interactive website is deployed by the service provider for researchers to conduct their studies in a secure manner and for data providers to authorize certain research requests.

Researchers

This includes the individuals or organizations who want to conduct research on multiple data providers' data sets. Researchers submit their requests to the service provider, which are then sent to the data providers for further processing.

As we use threshold homomorphic encryption to guarantee data and model security, in our proposed protocol, one public key corresponds to multiple secret keys, and different secret keys are distributed to different data providers and researchers. Furthermore, we assume that there exist at least one honest party and some semihonest adversaries that are capable of reading the internal information of the colluding parties while not deviating from the defined protocol [24].

Logistic Regression

Logistic regression is a classification algorithm that is widely used in medicine, including for disease diagnosis, clinical decision support, and risk assessment. Suppose a data set consists of pairs (x_i, y_i), for $i=1,...,N$, where x_i denotes a vector of input features $x_i=(x_i^1,...,x_i^d)$ and y_i is the class label. We then have:

In the sigmoid function $\sigma(x_i^T\beta)$, $\beta=(\beta_0,\beta_1,...,\beta_d)$ are the model parameters. By training a logistic regression model through minimization of the following cost function, we can obtain the optimal model parameters:

Homomorphic Encryption

Homomorphic encryption is a special type of encryption scheme that allows computations on ciphertexts without the need to access a secret key. Once the result of the computation is decrypted, it matches the result of the operations as if they were performed on the plaintext.

In our proposed protocol, we use a ring learning with errors (RLWE)–based, somewhat homomorphic encryption scheme, called Brakerski/Fan-Vercauteren (BFV) and which supports a limited number of additions and multiplications, to perform secure multiparty logistic regression [25,26]. The BFV scheme has some helpful properties for our protocol. First, it is more practical than the other 2 types of homomorphic encryption schemes, namely, partial and fully homomorphic encryption. More specifically, fully homomorphic encryption requires time-consuming bootstrapping to support an unlimited number of operations, whereas partial homomorphic encryption allows only addition or multiplication between ciphertexts. For example, the Paillier scheme only supports addition between ciphertexts, meaning that a ciphertext can only be multiplied by a plaintext, which results in massive transfer consumption if a large number of multiplications and the security of the plaintext are required [27]. Furthermore, some optimization techniques can be used to greatly improve the computation performance in the BFV scheme as long as we set the encryption parameters properly, such as number theoretic transform (NTT) and Chinese remainder theorem (CRT) batching [28]. Finally, the BFV scheme can be extended to threshold homomorphic encryption for secure multiparty computations.

The details of the threshold variant of the BFV scheme are described as follows. The security and noise analysis of the scheme are provided in Multimedia Appendix 1 [25,29]:

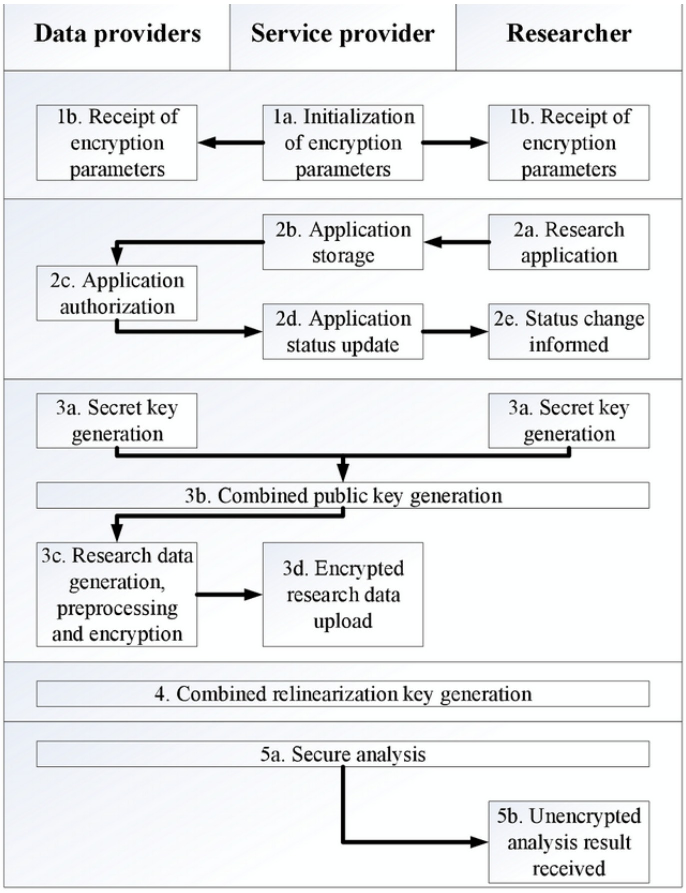
1. setup(1^λ): takes the security parameter λ as an input and returns the public parameterization param, including the degree of polynomial modulus n, the coefficient modulus q, the plaintext modulus t, and the (key, error) distribution (D1, D2).
2. THE.keygenSP(param): the service provider samples $a \leftarrow R_q$ and outputs it. Here, $R_q=Z_q[x]/(x^n+1)$ is the ciphertext space of param.
3. THE.keygenSkpk(param, a): each party p_i samples $s_i \leftarrow$ D1, $e_i \leftarrow$ D2, sets si as its secret key and outputs its public key $pk_i=[-(a \cdot s_i+e_i)]_q$. Let subscript $*_{co}$ denote the combined key. The combined public key pk_{co} among parties $p_1,\ldots,p_z$ is then computed as follows:
4. THE.keygenRelin(param, $s_1,\ldots,s_z$): the parties together with the service provider generate the combined relinearization key rlk_{co}. As the generation of the relinearization key is rather complicated, we will show the details of this step later.
5. THE.encrypt(m, pk_{co}): This takes a polynomial $m \in R_t$ as the input, where R_t is the plaintext space of the param. Let $pk_{co}=(pk_{co}(0), pk_{co}(1))$ and $\Delta= q/t$, and sample $u \leftarrow$ D1 and $(e_1, e_2) \leftarrow D_2$, then return:
6. THE.eval(C, rlk_{co}, $c_1,\ldots,c_c$): given a circuit C, a tuple of ciphertexts encrypted by the same public key, and the corresponding relinearization key, this outputs a ciphertext c_{out}. The procedure for homomorphic addition and multiplication is the same as that in the original single-key BFV scheme.
7. THE.decrypt(c, $s_1,\ldots,s_z$): given the ciphertext c=(c(0), c(1)) encrypted by pk_{co} and the corresponding secret keys, sample $(e_1, e_z) \leftarrow D_{smg}$. Here, the subscript $*_{smg}$ means that the variance of the noise distribution is much larger than that of the input ciphertext noise distribution to guarantee circuit privacy through smudging techniques [22]. The partial decryption shares are then computed as follows:

   These shares are sent to the party that requires the unencrypted result. The decryption result *m* is obtained by

Workflow of the Presented Protocol

The workflow of our proposed protocol consists of 5 major steps, as shown in Figure 2.

Figure 2. Workflow of the proposed protocol.



Initialization of Encryption Parameters

The service provider initializes the BFV homomorphic encryption parameters. These parameters should be carefully selected because they affect many aspects of the encryption scheme, such as operational performance, security level, multiplicative depth of the circuit, and space consumption. Two sets of parameters must be initialized by the service provider, one for the privacy-preserving logistic regression—$param1=(n1, q1, t1, D1_1, D1_2)$ and the other for the generation of the relinearization key in a secure manner—$param2=(n2, q2, t2, D2_1, D2_2)$. Once initialized, the 2 sets of parameters are sent to the data providers and researchers.

To make the encryption scheme practical, these parameters should meet the following criteria. First, the degree of polynomial modulus n must be a power of 2. Second, the coefficient modulus and the plaintext modulus must be either a prime P that satisfies P=1 (mod 2n) or a composite number that is a product of distinct primes, where every prime satisfies the above condition. After setting appropriate encryption parameters, NTT can be used to accelerate the multiplications between polynomials from o(n^2) to o($n$log$n$), whereas the adoption of CRT can improve the performance of the multiplications and additions of large integers, accelerating the multiplication and addition of the polynomials [30]. More importantly, we can apply CRT batching to greatly reduce space and computational consumption. Given a certain degree of polynomial modulus n, we can pack up to n values into one polynomial using CRT batching and apply the arithmetic operations to all the values within this polynomial in a single instruction, multiple data (SIMD) manner, whereas in a naive manner, we place a single value into one polynomial and apply operations to only one value.

Furthermore, to generate relinearization keys safely and correctly, the 2 sets of parameters must satisfy the following requirements: (1) their polynomial moduli must share the same degree and (2) the plaintext modulus in *param2* must be equal to the coefficient modulus in *param1*.

Research Application

The research application consists of several message transfers among the data providers, service providers, and researchers. First, a researcher visits the website deployed by the service provider and sets up a new research study. When the research begins, 3 settings must be confirmed by the researcher: first, the query condition used to obtain the research data; second, the list of data providers from which the researcher wishes to obtain the research data; finally, the settings of the secure logistic regression, including the variables to be used as features and the variable to be used as a class label and the settings of the maximum number of iterations, learning rate, and termination condition of the model training. This information is stored in the database of the service provider and sent to the corresponding data providers as a research request. After receiving the request, the data providers decide whether to authorize this research and send their decision to the service provider to inform the corresponding researcher about the authorization status.

Key Generation and Data Preparation

Once the data providers complete the research authorization, key generation is implemented by an interactive protocol among all parties, which comprises 2 steps—THE.keygenSP and THE.keygenSkpk. After this procedure, each party p_i holds its secret keys $s1_i$ and $s2_i$, whereas 2 corresponding public keys $pk1_{co}$ and $pk2_{co}$ are broadcasted among all parties. Here, the number in the symbol represents the set of parameters to which these keys belong.

The data preparation phase then begins, which is described as follows:

1. The data provider generates their own research data according to the query condition of the research. Next, all the floating-point numbers in the research data are scaled and rounded into integers because all the operations in the BFV scheme are integer based. Categorical features are encoded as integers if they are Boolean or ordered; otherwise, one-hot encoding is implemented.
2. The data provider encodes the research data by CRT batching. As mentioned before, we can pack multiple values into one polynomial and apply operations to them in an SIMD manner via CRT batching. This means that when given a data set with d features and N samples, one can pack them into d+1 polynomials (d features and 1 class label) as long as the degrees of the polynomial moduli are larger than N.
3. The data provider encrypts all the CRT-batched polynomials using the combined public key $pk1_{co}$. After all the plaintext polynomials are encrypted, they are sent to the service provider.

Relinearization Key Generation

After data preparation, the researcher, and all involved data providers together with the service provider generate the combined relinearization key. The relinearization step is not necessary for the correctness of homomorphic multiplication but is essential in our threshold-variant BFV scheme. By performing relinearization after every homomorphic multiplication, the size of the ciphertext can be strictly kept at 2, which simplifies decryption.

The relinearization key generation procedure is illustrated next. We denote the number of parties by z. Suppose the coefficient modulus in *param1* is a product of k distinct primes, whereas each party p_i holds 2 secret keys $s1_i$ and $s2_i$ from *param1* and *param2*, respectively. Given a combined public key $pk2_{co}$ from *param2*, the following is observed:

1. Each party p_i performs THE.encrypt($s1_i$, $pk2_{co}$) and outputs k ciphertexts, of which the plaintext modulus is a group of primes whose product is the coefficient modulus in param1. The ciphertexts of secret key $c_j(s1_i)$ (j=1,...,k) are then sent to the service provider.
2. The service provider computes the ciphertexts of the combined secret key $c_j(s1_{co})$ (j=1,...,k) and sends them to the data provider and researcher:
3. Each party p_i computes the ciphertexts of the product of the combined secret key and its secret key from *param1* as follows and sends the result to the service provider:

   Here, $c_j(0)$ (j=1,...,k) are the ciphertexts of 0, which contain sufficiently large noise to guarantee function privacy [31].
4. The service provider computes the ciphertexts of the square of the combined secret key $c_j(s1_{co}2)$ (j=1,...,k) as follows:

   Having encrypted the combined secret key and its square, the service provider defines the decomposition bit count T and the size of the relinearization key $L = \log_2(q1)/T$, samples $a_0 \sim a_L \leftarrow R_{q1}$, whereas each party p_i samples $e_{i0} \sim e_{iL} \leftarrow D1_2$, performs THE.encrypt($e_{i0} \sim e_{iL}$, $pk2_{co}$) and sends these ciphertexts $c_j(e_{i0} \sim e_{iL})$ (j=1,...,k) to the service provider. After receiving encrypted noise, the service provider computes the following:

   The encrypted combined relinearization key is then generated as follows: all parties perform THE.decrypt($c_j(rlk_{co})$, $s2_1$,...,$s2_z$) and finally return the plaintext combined relinearization key rlk_{co}. Compared with the combined relinearization key generation procedure presented in the study by Mouchet et al [22], our method involves more transfer consumption but much less noise, which grows only linearly with the number of parties

Privacy-Preserving Model Training and Evaluation

Secure logistic regression model training begins once all the encrypted research data and the combined relinearization key are sent to the service provider. We choose the gradient descent algorithm to train the model with homomorphically encrypted data because we can implement the algorithm using only addition and multiplication, which all fully and somewhat homomorphic encryption schemes naturally have, whereas despite its faster convergence, Newton method requires matrix inversion, which may have a very high time cost under the homomorphic encryption computation [32].

After choosing the proper training method, another major problem is the evaluation of the sigmoid function $\sigma(x^T\beta)$, because the BFV scheme can only be used to evaluate polynomial functions. Instead of simply using the Taylor polynomial to approximate the sigmoid function, we use the degree-3 least squares approximation of the sigmoid function over the interval (−8, 8), as the former has a much larger error as $|x^T\beta|$ increases, whereas the latter only has a small error as long as $x^T\beta$ is within the interval [13]. The least squares approximation polynomial is:

As the BFV scheme is based on integers, we apply scaling factor (SF) to scale up the floating-point number $x^T\beta$ into the integer $x^T\beta \times SF$. In our privacy-preserving logistic regression protocol, we set SF=1000, which is a trade-off between approximation accuracy and performance. Specifically, if we set SF smaller, the approximation accuracy will decrease; if we set SF larger, 2 or more polynomials may be required to represent a set of values, or larger encryption parameters may be required to maintain the same multiplicative depth for a given security level, both of which result in larger space and computational resource consumption. This SF also scales up the approximation interval from (−8, 8) to (−8000, 8000), scaling the degree-1 and degree-3 coefficients to 1/1000 and $1/1000^3$, respectively, of the original value. Finally, the least squares approximation function is integerized to be compatible with the homomorphic encryption computation:

The integerized function output is then transformed into an original function:

We now describe the detailed process of secure logistic regression. Before training begins, the involved data providers divide their own research data into 10 folds from *Fold1~Fold10* for 10-fold cross-validation and then encode the information into a vector. For example, a data set x containing 20 samples is divided as follows:

Fold1 ~ (x_1, x_6), *Fold2* ~ (x_2, x_{17}), *Fold3* ~ (x_3, x_{13}), *Fold4* ~ (x_4, x_{10}), *Fold5* ~ (x_5, x_{20}), *Fold6* ~ (x_7, x_{16}), *Fold7* ~ (x_8, x_{14}), *Fold8* ~ (x_9, x_{11}), *Fold9* ~ (x_{12}, x_{18}), *Fold10* ~ (x_{15}, x_{19})

Next, the information is encoded into a vector of values (1, 2, 3, 4, 5, 1, 6, 7, 8, 4, 8, 9, 3, 7, 10, 6, 2, 9, 10, 5). The vector can be viewed as a special column of research data, although this column is not used in the computation of the approximation sigmoid function.

When all the data providers finish dividing their research data, they send these vectors to the service provider. As these vectors do not contain any sensitive information, they do not need to be further encoded into CRT-batched polynomials and encrypted.

After all preparations are completed, the model training begins, as shown in Textboxes 1-3. In Textbox 1, we use minibatch gradient descent instead of batch gradient descent because the former converges faster, and we can make full use of CRT batching by simultaneously training 10 models for 10-fold cross-validation, which vastly reduces the time cost of model training. Specifically, for each iteration, the researcher assigns the sets of parameters to the research samples according to the number of iterations and the fold to which these samples belong, which means that in one iteration, a one-to-one correspondence exists between the 10 sets of parameters and the 10 folds of research data. Once the gradient ciphertexts are computed, all data providers will mask them via randomly generated encrypted noises (Textbox 3). The masked gradients are then decrypted, and only the researcher can obtain the plaintext result. As the researcher only knows the sum of noises for each fold, the correct overall gradients are finally obtained to update the model parameters of the researcher without revealing the gradient of any single sample.

Textbox 1. Privacy-preserving logistic regression model training.

Input: *epoch* (# of iterations), α (learning rate), ε (step tolerance), $c(x)=\{c(x^1),...,c(x^d),c(y)\}$ (encrypted research data), x^{d+1} (vector describing how data providers divide their research data), b (# of samples in one fold), z (# of parties), $s1_1 \sim s1_z$ (secret keys), $pk1_{co}$ (combined public key), $\beta(1) \sim \beta(10)$ (model parameters initialized by researcher where each $\beta(i)=(\beta(i)^0, \beta(i)^1,...,\beta(i)^d)$)

Output: $\beta_{new}(1) \sim \beta_{new}(10)$ (trained model parameters)

Researcher does:

```
For iter=1 to epoch / 9
    $\beta_{old}(1) \sim \beta_{old}(10) \leftarrow \beta(1) \sim \beta(10)$
    For cv=1 to 9
        $B \leftarrow$ empty vector
        For-each element $i$ in $x^{d+1}$
            $B$.push_back($\beta((i+cv-1) \bmod 10+1)$)
        End for-each
        $B' \leftarrow$ CRT-batchingEncode($B$) // $B'=\{B'^0,\dots,B'^d\}$
        $c(B^0) \sim c(B^d) \leftarrow$ THE.encrypt($B'$, $pk1_{co}$)
        Wait for encrypted gradient calculation $c(gra^0) \sim c(gra^d)$ // See (Textbox 2) for details
        Wait for securely decryption of encrypted gradients $gra(1) \sim gra(10)$ // See (Textbox 3) for details
        $\beta(1) \sim \beta(10) -= (gra(1) \sim gra(10))\times\alpha \div b$
    End for
    $\beta_{new}(1) \sim \beta_{new}(10) \leftarrow \beta(1) \sim \beta(10)$
    If ($||\beta_{new}-\beta_{old}|| \div ||\beta_{new}||<\varepsilon$) then
        return $\beta_{new}(1) \sim \beta_{new}(10)$
    End if
End for
```

Textbox 2. Encrypted gradient calculation.

Input: $c(B^0) \sim c(B^d)$, $c(x)$ // See details in (Textbox 1)

Output: $c(gra^0) \sim c(gra^d)$ (encrypted gradients)

Service provider does:

```
$c(x^T\beta) \leftarrow c(B^0)+c(B^1)\times c(x^1)+...+c(B^d)\times c(x^d)$
$c(G) \leftarrow G_3(c(x^T\beta))$ // $G_3$ is an integerized sigmoid function
$c(gra^0) \sim c(gra^d) \leftarrow [c(G)-627743311836\times c(y)]\times[c(x^0) \sim c(x^d)]$ // Here, $c(x^0)=1$
```

Textbox 3. Secure decryption of encrypted gradients.

```
Input: $x^{d+1}$, $cv$, $c(gra^0) \sim c(gra^d)$, $s1_1 \sim s1_z$, $pk1_{co}$ // See details in (Textbox 1)
Output: $gra(1) \sim gra(10)$ (unencrypted gradients)
All data providers do:
1: $e^0 \sim e^d \leftarrow$ random noise vectors whose size equals $x^{d+1}$
2: $E(1) \sim E(10) \leftarrow$ zero vectors whose size equals $d+1$
3: For $i=1$ to size($x^{d+1}$)
4:     For $j=1$ to $d+1$
5:         $E((x^{d+1}(i)+cv-1) \bmod 10+1)(j) \mathrel{+}= e^{j-1}(i)$
6:     End for
7: End for // $E(1) \sim E(10)$ are sent to the researcher
8: $e' \leftarrow$ CRT-batchingEncode($e^0 \sim e^d$)
9: $c(e^0) \sim c(e^d) \leftarrow$ THE.encrypt($e'$, $pk1_{co}$) // $c(e^0) \sim c(e^d)$ are sent to the service provider
Service provider does:
10:$c'(gra^0) \sim c'(gra^d) \leftarrow c(gra^0) \sim c(gra^d)+c(e^0) \sim c(e^d)$
All parties do:
11:$gra'^0 \sim gra'^d \leftarrow$ THE.decrypt($c'(gra^0) \sim c'(gra^d)$, $s1_1 \sim s1_z$) // To ensure only the researcher obtains the plaintext result, data providers' and researcher's partial decryption shares are added at the service provider and the researcher, respectively.
Researcher does:
12:$gra(1) \sim gra(10) \leftarrow$ zero vectors whose size equals $d+1$
13:$gra''^0 \sim gra''^d \leftarrow$ CRT-batchingDecode($gra'^0 \sim gra'^d$) // Decoding result is vectors whose size equals $x^{d+1}$.
14:For $i=1$ to size($x^{d+1}$)
15:    For $j=1$ to $d+1$
16:        $gra((x^{d+1}(i)+cv-1) \bmod 10+1)(j) \mathrel{+}= gra''^{j-1}(i)$
17:    End for
18:End for
19:$gra(1) \sim gra(10) \mathrel{-}= E(1) \sim E(10)$
```

Once the model training is completed, all involved data providers encode their own research data for each fold into CRT-batched polynomials whose slots are randomly chosen to contain samples. In the meantime, the data providers also generate vectors containing information about whether a certain slot contains a sample and encode them into CRT-batched polynomials. For instance, for a CRT-batched polynomial containing samples in slots (1, 6, 8), the vector should be (1, 0, 0, 0, 0, 1, 0, 1). These polynomials are then encrypted by $pk1_{co}$ and sent to the service provider.

When all the aforementioned preparations are completed, the model evaluation starts, as shown in Textboxes 4-6. In Textbox 5, lines 3-5, all data providers mask the encrypted predictive values. Here, the noise generation should meet 2 criteria, whereas the noise generation in Textbox 3 line 1 has no special limitations as long as the error is random and sufficiently large to mask the true values. First, in the empty slots, we sample noise from a uniform distribution whose upper and lower bounds are the minimum and maximum values of the integerized approximation sigmoid function G_3. Second, in the slots containing samples, we sample noise from a uniform distribution (−1569358279, 1569358279) whose corresponding values are (−0.005, 0.005) in the scaled down plaintext. In Textbox 6, lines 1-3, all data providers perform another masking; this time, the noise generation is exactly the same as in Textbox 3 line 1.

Textbox 4. Model evaluation.

```
Input: $c(1)(x, y) \sim c(10)(x, y)$ (10 encrypted folds of research data), $c(1)(x^{d+1}) \sim c(10)(x^{d+1})$ (encrypted vectors indicating whether a certain slot contains a sample), $\beta(1) \sim \beta(10)$ (trained sets of parameters), $s1_1 \sim s1_z$ (secret keys), $pk1_{co}$ (combined public key)
Output: $TP, FP, TN, FN$ (number of true positives, false positives, true negatives, and false negatives, respectively, under different predictive value thresholds)
Researcher does:
For $FD$=1 to 10
    $c(\beta^0) \sim c(\beta^d) \leftarrow$ THE.encrypt($\beta(FD), pk1_{co}$) // $c(\beta^0) \sim c(\beta^d)$ are sent to the service provider
    Wait for masked predictive values $\sigma$ // See (Textbox 5) for details
    For $V=\min(G_3) : (\max(G_3)-\min(G_3))/100 : \max(G_3)$
        $X \leftarrow$ empty vector
        For-each predictive value $\sigma_i$ in $\sigma$
            $X$.push_back(if($\sigma_i \geq V$))
        End for-each
        $X' \leftarrow$ CRT-batchingEncode($X$)
        $c(TP), c(FP), c(TN), c(FN) \leftarrow c(FD)(y) \times X' \times c(FD)(x^{d+1}), (1-c(FD)(y)) \times X' \times c(FD)(x^{d+1}), (1-c(FD)(y)) \times (1-X') \times c(FD)(x^{d+1}), c(FD)(y) \times (1-X') \times c(FD)(x^{d+1})$ // These 4 ciphertexts are sent to the service provider
        Wait for masked model evaluation results $TP', FP', TN', FN'$ // See (Textbox 6) for details
        $TP'', FP'', TN'', FN'' \leftarrow$ CRT-batchingDecode($TP', FP', TN', FN'$)
        $TP, FP, TN, FN \leftarrow TP''-\mathrm{sum}(e_{TP}), FP''-\mathrm{sum}(e_{FP}), TN''-\mathrm{sum}(e_{TN}), FN''-\mathrm{sum}(e_{FN})$
        output $TP, FP, TN, FN$ // under fold $FD$ and predictive value threshold $V$
    End for
End for
```

Textbox 5. Calculation of masked predictive values.

```
Input: $c(1)(x) \sim c(10)(x), x^{d+1}(1) \sim x^{d+1}(10), c(\beta^0) \sim c(\beta^d), FD, pk1_{co}, s1_1 \sim s1_z$ // See details in (Textbox 4)
Output: $\sigma$ (masked predictive values)
Service provider does:
    $c(x^T\beta) \leftarrow c(\beta^0)+c(\beta^1) \times c(FD)(x^1)+...+c(\beta^d) \times c(FD)(x^d)$
    $c(G) \leftarrow G_3(c(x^T\beta))$ // $G_3$ is an integerized sigmoid function
All data providers do:
    $e \leftarrow$ random noise vectors whose size equals $x^{d+1}(FD)$
    $e' \leftarrow$ CRT-batchingEncode($e$)
    $c(e') \leftarrow$ THE.encrypt($e', pk1_{co}$) // $c(e')$ are sent to the service provider
Service provider does:
    $c'(G) \leftarrow c(G)+c(e')$
All parties do:
    $\sigma \leftarrow$ CRT-batchingDecode(THE.decrypt($c'(G), s1_1 \sim s1_z$)) // The same as in (Textbox 3), only the researcher obtains the plaintext result
```

Textbox 6. Calculation of masked model evaluation results.

```
Input: $c(1)(x^{d+1}) \sim c(10)(x^{d+1})$, $FD$, $pk1_{co}$, $c(TP)$, $c(FP)$, $c(TN)$, $c(FN)$, $s1_1 \sim s1_z$ // See details in (Textbox 4)
Output: $TP'$, $FP'$, $TN'$, $FN'$ (masked model evaluation results)
All data providers do:
    $e_{TP}, e_{FP}, e_{TN}, e_{FN} \leftarrow$ random noise vectors whose size equals to $c(FD)(x^{d+1})$ // The sums of noises $sum(e_{TP})$, $sum(e_{FP})$, $sum(e_{TN})$, $sum(e_{FN})$ are sent to the researcher
    $e'_{TP}, e'_{FP}, e'_{TN}, e'_{FN} \leftarrow$ CRT-batchingEncode($e_{TP}, e_{FP}, e_{TN}, e_{FN}$)
    $c(e'_{TP}), c(e'_{FP}), c(e'_{TN}), c(e'_{FN}) \leftarrow$ THE.encrypt(($e'_{TP}, e'_{FP}, e'_{TN}, e'_{FN}$), $pk1_{co}$) // These encrypted noises are sent to the service provider
Service provider does:
    $c'(TP), c'(FP), c'(TN), c'(FN) \leftarrow c(TP)+c(e_{TP}), c(FP)+c(e_{FP}), c(TN)+c(e_{TN}), c(FN)+c(e_{FN})$
All parties do:
    $TP', FP', TN', FN' \leftarrow$ THE.decrypt(($c'(TP), c'(FP), c'(TN), c'(FN)$), $s1_1 \sim s1_z$) // The same as in (Textbox 3), only the researcher obtains the plaintext result
```

Once the model evaluation ends, the researcher obtains the number of true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) for the 10 folds and different predictive value thresholds, which should be sufficient to evaluate the trained model via 10-fold cross-validation.

Results

Overview

In this section, we consider the following aspects to assess the performance of our proposed multicenter secure logistic regression protocol: (1) Security analysis: security of sensitive research data and learned model; (2) accuracy loss: the loss in accuracy during the model training and evaluation with respect to the nonsecure method with real medical data; (3) model training and evaluation time: the time needed to perform 10-fold cross-validation with real medical data; and (4) scalability: how the model training and evaluation time increases as the size of the data increases in the synthetic data set.

The biomedical data sets used for the experiments are shown in Table 1 [33,34]. For the breast cancer data set, we eliminate missing samples, use all the attributes except breast-quad, and assume that the data set is provided by 1 data provider. For the surveillance, epidemiology, and end results colorectal cancer data set, we choose a portion of the samples and use 5-year survival status as the label. Moreover, all the attributes, except the registry, are used, and we assume that the data set is provided by 3 different data providers. More details about these 2 data sets are provided in Multimedia Appendix 1. We use 10-fold cross-validation, which partitions the data sets into 10 folds of approximately equal size by stratified sampling to ensure that the positive/negative ratio of each fold is approximately equal. Each time, 9 folds are used as the training set and the remaining fold is used as the test set. In addition, we assume that during model training, all data ciphertexts share the same data division vector so that the gradient ciphertexts can be summed to reduce the size of transferred data in Textbox 3 line 11.

Table 1. Description of the data sets.

Data sets	SEER[a] CRC[b] data [31]	UCI[c] breast cancer [32]
Samples, n	49152	277
Attributes, n	10	9
Size of ciphertexts, MB	60.0	18.0

[a] SEER: surveillance, epidemiology, and end results.

[b] CRC: colorectal cancer.

[c] UCI: unique client identifier.

To set the homomorphic encryption parameters, we select the following parameters to guarantee sufficient security, as shown in Table 2. Our values for the polynomial modulus, coefficient modulus, and security level match the most recent homomorphic encryption security standards proposed by the Homomorphic-Encryption.org group [35]. The degree of polynomial modulus n is a power of 2, whereas the coefficient moduli in *param1* and *param2* are products of 8 and 5 distinct primes, respectively, where every prime P is at most 60 bits long and satisfies $P=1 \pmod{2n}$, which makes the NTT accessible. The plaintext modulus in *param1* also satisfies $t1=1 \pmod{2n}$, allowing for the implementation of CRT batching.

Table 2. Select parameters for Brakerski/Fan-Vercauteren homomorphic encryption.

Parameters	*param1*	*param2*
Polynomial modulus	16,384	16,384
Coefficient modulus	438-bit integer	300-bit integer
Plaintext modulus	1125899904679937	Coefficient modulus of *param1*
Key distribution	Uniform distribution {−1, 0, 1}	Uniform distribution {−1, 0, 1}
Error distribution	Discrete Gaussian distribution, with σ=3.2	Discrete Gaussian distribution with σ=3.2
Security level	128-bit	192-bit

To simulate a real-world scenario, we place the data providers, the researcher, and the service provider on different machines. For the data providers and the researcher, we use PCs with a 2.2-GHz Intel Core i7-8750H processor and 16.0 GB RAM (Windows 10 Enterprise). For the service provider, we use a server with a 2.3 GHz Intel Xeon Gold 6140 processor and 128.0 GB RAM (Linux 3.10.0). The secure logistic regression protocol is implemented in C++ using Microsoft SEAL v3.0 and is publicly available at GitHub [36], where we made some modifications to support the threshold-variant BFV scheme [37]. All PCs have an internet connection of 100 Mbps bandwidth.

Security Analysis

In our protocol, security means that corrupted parties will not be able to obtain sensitive data or learned models from honest parties. Here, we show the security of our protocol from the following 2 aspects: (1) honest parties' secret keys will not be obtained by the corrupted parties so that no ciphertext will be decrypted illegally, including the encrypted data, model parameters, and any other intermediate results and (2) if the researcher is an adversary, he or she cannot obtain any meaningful information about honest parties' individuals from the unencrypted intermediate results.

Security of Secret Keys

To demonstrate the security of the secret keys, we use the simulation paradigm described in the study by Goldreich [38], that is, for all adversaries, there exists a simulator program S that, when provided only with the adversaries' input and output, can simulate the adversaries' view in the protocol, and the simulated view is computationally indistinguishable from the real view. Suppose there are z parties. Let A denote the adversaries, defined as a subset of at most $z-1$ corrupted parties, and H denote the honest parties.

Combined Public Key Generation

In the generation of the combined public key, S can simulate the adversaries' view of public key shares (pk_1, pk_2,..., pk_z) by randomizing these shares under 2 constraints: (1) the simulated shares must sum to $pk_{co}(0)$ and (2) the adversary shares must be equal to the real shares. S can compute this sharing as follows:

When $|H|>1$, there is no efficient algorithm that can distinguish between the simulated and real shares in H because of the decision-RLWE problem. When $|H|=1$, S computes the real shares of the honest party. However, because both s_i and e_i are private inputs from party p_i, the adversaries cannot find the secret key of the honest party because of the search-RLWE problem.

Decryption

Given the ciphertext $c=(c(0), c(1))$, during the decryption process, S can simulate the adversaries' view of the decryption shares (μ_1, μ_2,..., μ_z) by randomizing these shares under 2 constraints: (1) the simulated shares must sum to $\mu-c(0)$ and (2) the adversary shares must be equal to the real shares:

When considering the distribution of the simulated and real views alone, the RLWE assumption is sufficient to ensure the security of secret keys of H if the researcher is uncorrupted. However, if the researcher becomes an adversary, they can extract the noise of c as follows:

where e' is the noise of c, which should be unknown to the researcher; otherwise, the RLWE assumption will be broken and the secret keys of the honest parties may be exposed to the researcher. Let var^2_c denote the variance of a centered Gaussian distribution that e follows and var^2_{smg} denote the variance of D_{smg}, which is used to generate e_i. Thus, as long as the ratio var^2_c/var^2_{smg} is negligible, the following 2 distributions are statistically indistinguishable, which means that e' is unknown to the researcher and that the researcher cannot obtain H's secret keys:

Unencrypted Intermediate Results

First, during model training, all data providers apply one-time-use noise to mask the encrypted gradient before decryption, meaning that even if only one data owner is honest, it will not lead to the disclosure of the gradients of the individuals.

Second, during model evaluation, the researcher will inevitably obtain CRT-batched polynomials containing the predictive values for each sample. Given a masked predictive value $\sigma_i \in$

(V_j, V_{j+1}), the probability of recovering the research data is computed as follows:

Here, N_j is the number of samples whose predictive value belongs to (V_j, V_{j+1}), N_{ej} is the number of empty slots whose value belongs to (V_j, V_{j+1}), and N_{pi} is the number of all possible combinations of feature values whose predictive value belongs to (σ_i−1569358279, σ_i+1569358279). Therefore, as long as either of these 2 terms is sufficiently small, it is impossible for the researcher to recover the feature values.

Furthermore, because the encrypted (TP, FP, TN, and FN) information of samples under different predictive value thresholds is also masked by all data providers before being sent to the researcher, the researcher cannot obtain the label of any specific sample.

Accuracy Loss

In Table 3, we demonstrate the accuracy of our protocol by comparing the area under the curve between the nonsecure logistic regression and our secure logistic regression, where the former uses the standard sigmoid function and both have the same hyperparameters (learning rate α=.1, 45 iterations). Compared with that of the nonsecure protocol, a relatively small loss of accuracy was observed in our protocol, which was not statistically significant (the smallest P=.09). The average receiver operating characteristic curves from the 10-fold cross-validation are plotted in Figure 3.

Table 3. Accuracy comparison between nonsecure and proposed secure logistic regressions.

Data sets	SEER[a] CRC[b] data	Breast cancer
AUC[c] (nonsecure)	0.703 (0.008)	0.728 (0.156)
AUC (our protocol)	0.696 (0.008)	0.717 (0.164)
P value (AUC)	.09	.88
Accuracy (nonsecure)	0.620 (0.013)	0.664 (0.149)
Accuracy (our protocol)	0.612 (0.013)	0.632 (0.155)
P value (accuracy)	.18	.64
F_1[d] (nonsecure)	0.654 (0.012)	0.508 (0.198)
F_1 (our protocol)	0.649 (0.012)	0.505 (0.240)
P value (F_1)	.42	.97

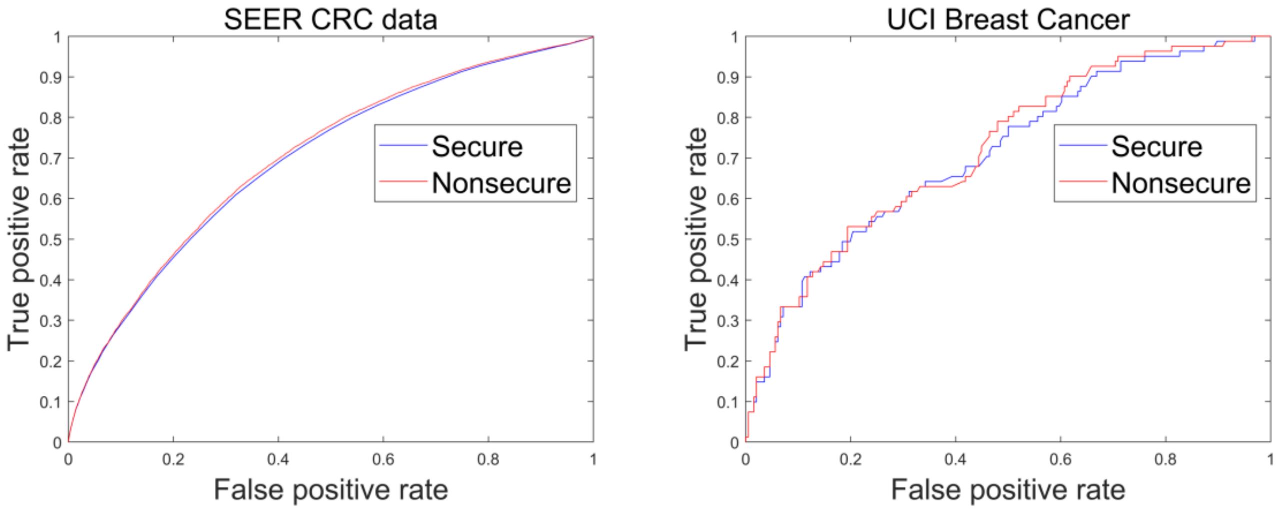
[a] SEER: surveillance, epidemiology, and end results.

[b] CRC: colorectal cancer.

[c] AUC: area under the curve.

[d] F_1: the harmonic mean of the precision and recall.

Figure 3. Average receiver operating characteristic curves of nonsecure and proposed secure logistic regressions. CRC: colorectal cancer; ROC: receiver operating characteristic; SEER: surveillance, epidemiology, and end results; UCI: University of California, Irvine.



Furthermore, in Table 4, we test the relationships between the learning rate and the convergence of the nonsecure and secure logistic regressions. Although our protocol's model training will be fully spoiled because of the limited valid input interval for the approximation sigmoid function when the learning rate becomes too large, our protocol has a slightly broader range of learning rate selection than the nonsecure protocol.

Table 4. ||βnew–βold|| ÷ ||βnew|| after 99 iterations (surveillance, epidemiology, and end results colorectal cancer data).

Learning rate	0.1	0.2	0.3	0.4
Nonsecure	0.056	0.046	0.302	0.347
Our protocol	0.061	0.052	0.047	—[a]

[a] Fail to convergence.

Model Training and Evaluation Time

We show the time consumption of the 10-fold cross-validation for the 2 different data sets in Table 5.

Here, we compare our protocol with the SecureLR protocol by Jiang et al [19], which is also optimized with NTT and CRT batching but evaluated on only 1 PC. As shown in their experiments, SecureLR can train only 1 model at a time and requires 44.9 seconds per iteration over a data set with a ciphertext size of 5.0 M. In comparison, our protocol can train 10 models simultaneously and perform each iteration much faster (on a data set with a ciphertext size of 60.0 M in less than 10 seconds per iteration). Moreover, our protocol supports secure model evaluation with reasonable time consumption.

Table 5. Time consumption of the proposed protocol.

Data sets	Iterations, n	Training time	Time per iteration (seconds)	Evaluation time
SEER[a] CRC[b] data	45	7 min 29 seconds	9.98	20 min 27 seconds
UCI[c] breast cancer	45	4 min 24 seconds	5.87	14 min 28 seconds

[a] SEER: surveillance, epidemiology, and end results.

[b] CRC: colorectal cancer.

[c] UCI: unique client identifier.

Scalability Evaluation

To test our protocol's scalability, we use a synthetic data set with different numbers of data providers and features, as shown in Tables 6 and 7. Given a certain number of features *d*, for the sake of simplicity, we suppose that every data provider encrypts (*d*+1) polynomials. As the number of data providers increases, the computation times of both the model training and evaluation increase proportionally, whereas there is no increase in the transfer time of the model training because the size of the transferred data (encrypted parameters and gradients) is only related to the number of features. Similarly, because there is no relationship between the number of data providers and the transfer of the encrypted (TP, FP, TN, and FN), the transfer time of the model evaluation increases very less. As the number of features increases, the computation and transfer times of the model training increase proportionally, whereas the computation and transfer times of the model evaluation increase only slightly because the majority of the model evaluation involves the computation of (TP, FP, TN, and FN) information under different predictive value thresholds, which is not related to the number of features.

Table 6. Scalability of the proposed protocol for different numbers of data providers (9 features).

Data providers, n	Size of ciphertexts, MB	Iterations, n	Training time (computation)	Training time (transfer)	Evaluation time (computation)	Evaluation time (transfer)
3	60.0	45	4 min 16 seconds	3 min 13 seconds	9 min 54 seconds	10 min 33 seconds
5	100.0	45	6 min 26 seconds	3 min 13 seconds	15 min 24 seconds	10 min 39 seconds
10	200.0	45	12 min 45 seconds	3 min 12 seconds	30 min 42 seconds	10 min 51 seconds
15	300.0	45	19 min 5 seconds	3 min 13 seconds	45 min 54 seconds	11 min 3 seconds
20	400.0	45	25 min 52 seconds	3 min 13 seconds	61 min 13 seconds	11 min 17 seconds

Table 7. Scalability of the proposed protocol for different numbers of features (3 data providers).

Features, n	Size of ciphertexts, MB	Iterations, n	Training time (computation)	Training time (transfer)	Evaluation time (computation)	Evaluation time (transfer)
3	60.0	45	4 min 16 seconds	3 min 13 seconds	9 min 54 seconds	10 min 33 seconds
5	100.0	45	8 min 30 seconds	6 min 23 seconds	10 min 22 seconds	10 min 53 seconds
10	200.0	45	12 min 48 seconds	9 min 37 seconds	10 min 47 seconds	11 min 13 seconds
15	300.0	45	16 min 54 seconds	12 min 50 seconds	11 min 16 seconds	11 min 32 seconds
20	400.0	45	21 min 13 seconds	16 min 10 seconds	11 min 40 seconds	11 min 53 seconds

Discussion

Principal Findings

As researchers cannot obtain unencrypted research data, they may have difficulty choosing the proper hyperparameters, especially the learning rate. Despite a slightly broader range of learning rate selection, the setting of the learning rate is still very important in our privacy-preserving multicenter logistic regression protocol because compared with the nonsecure protocol, our protocol still has a considerable time cost. In our proposed protocol, interactions exist among the service provider, the data providers, and the researcher, allowing the researcher to obtain the plaintext model parameters in every iteration. As a result, the researcher can easily judge whether the hyperparameters are set properly according to the trend of the model parameters. Moreover, the researcher can halt the model training in the early stages, which results in less waste of computational resources. However, to implement the web-based protocol, clients must be installed on all the data providers' and researchers' machines, which must be kept online during the entire process of model training and model evaluation, leading to an additional consumption of network bandwidth.

There is a trade-off between computation and transfer consumption in our protocol. Although some solutions use fully homomorphic encryption to avoid decryption during model training [14,15], our proposed protocol uses somewhat homomorphic encryption for several reasons. First, to support an unlimited number of operations, a bootstrapping process is required, which is very time consuming. More time is consumed in threshold homomorphic encryption because we must select larger encryption parameters because there is not only greater noise in the combined public and relinearization keys but also greater smudging noise during decryption. Second, to avoid decryption, fixed-point arithmetic operations without a rounding process are required. Bonte and Vercauteren [14] use nonintegral base nonadjacent form with window size ω to encode a real number as a polynomial, which may affect the use of CRT batching (the most important optimization technique in our protocol), whereas Chen et al [15] use the Cheon-Kim-Kim-Song (CKKS) [39] scheme, which is also based on RLWE and naturally supports floating-point approximate arithmetic operations. However, in the CKKS scheme, the decryption result contains noise, meaning that in the threshold variant of the CKKS scheme, we must set a very high value for the encryption parameter *scale* to avoid destruction of the plaintext by the smudging noise, which greatly reduces the multiplicative depth of the circuit.

Limitations

Our proposed protocol has a few limitations. First, to make the privacy-preserving logistic regression realistic, this protocol requires a high-speed and stable network. Second, as the BFV scheme is based on integers, before encryption, all floating-point numbers must be scaled up and rounded to integers. A larger SF can support a higher level of precision but will also result in higher computation and storage costs for a given security level. Third, in a real-world scenario, a single patient may have multiple medical records across different data providers, which rarely occurs when data providers are far apart but is not uncommon when data providers are located in the same region (eg, a city). Therefore, in the latter case, further research on privacy-preserving identification and deduplication is required to ensure that there are no duplicate medical records to affect the analysis results. Furthermore, this study mainly focuses on technical issues and thus does not delve into matters related to ethics and law, which are also very important in multiparty medical research.

Conclusions

In this paper, we propose the first privacy-preserving multiparty logistic regression model training and evaluation protocol based on threshold homomorphic encryption. We conduct experiments in simulated real-life scenarios, and the results demonstrate that the proposed protocol is practical for real-world use. We believe that our work can help medical institutions eliminate privacy leakage concerns during data sharing, promote multicenter medical research, and thus improve the use of medical data to some extent.

In the future, we will extend our tools to be more practical. As the BFV homomorphic encryption scheme does not have indistinguishability under chosen ciphertext attack security, additional security technology, such as hashing, should be integrated into the tools to prevent malicious attackers from tampering with the ciphertexts. More privacy-preserving statistics and machine learning methods will be added to our tools to facilitate considerably enhance flexibility in secure multicenter research. Furthermore, we will improve the efficiency of our tools using graphics processing unit or field programmable gate array acceleration.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (under Grant 81771936 and 81801796), the Major Scientific Project of Zhejiang Laboratory (under Grant 2018DG0ZX01), the National Key Research and Development Program of China (under Grant 2018YFC0116901), and the Fundamental Research Funds for the Central Universities, China (No. 2020QNA5031).

Authors' Contributions

The study concept and design were given by YL and TZ. Implementation and experiments of the study were carried out by YL. Drafting of the manuscript was carried out by YL and YT. Discussion, critical revision, and final approval of the version to be published were performed by JL, SZ, YT, TZ, and YL.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of used biomedical data, details of Brakerski/Fan-Vercauteren (BFV) threshold homomorphic encryption, security analysis, and noise analysis.

[DOCX File , 38 KB - jmir_v22i12e22555_app1.docx]

References

1. Chung KC, Song JW, WRIST Study Group. A guide to organizing a multicenter clinical trial. Plast Reconstr Surg 2010 Aug;126(2):515-523 [FREE Full text] [doi: 10.1097/PRS.0b013e3181df64fa] [Medline: 20375760]
2. Downey AS, Olson S. Sharing Clinical Research Data: Workshop Summary. Washington, DC: National Academies Press; 2013.
3. Dirnagl U, Fisher M. International, multicenter randomized preclinical trials in translational stroke research: it's time to act. J Cereb Blood Flow Metab 2012 Jun;32(6):933-935 [FREE Full text] [doi: 10.1038/jcbfm.2012.51] [Medline: 22510602]
4. How are Healthcare Data Breach Victims Affected by Attacks? HealthITSecurity. URL: https://healthitsecurity.com/news/how-are-healthcare-data-breach-victims-affected-by-attacks [accessed 2020-06-17]
5. Skinner ER, Barnett GO, Singer DE, Mulley AG, Lew RA, Stickler SA, et al. The use of logistic regression in diagnostic and prognostic prediction in a medical intensive care unit. Proc Annu Symp Comput Appl Med Care 1980 Nov;1:222-227 [FREE Full text]
6. Abdolmaleki P, Yarmohammadi M, Gity M. Comparison of logistic regression and neural network models in predicting the outcome of biopsy in breast cancer from MRI findings. Iran J Radiat Res 2004;1(4):217-228 [FREE Full text]
7. Wu Y, Jiang X, Kim J, Ohno-Machado L. Grid Binary LOgistic REgression (GLORE): building shared models without sharing data. J Am Med Inform Assoc 2012;19(5):758-764 [FREE Full text] [doi: 10.1136/amiajnl-2012-000862] [Medline: 22511014]
8. Wang S, Jiang X, Wu Y, Cui L, Cheng S, Ohno-Machado L. EXpectation Propagation LOgistic REgRession (EXPLORER): distributed privacy-preserving online model learning. J Biomed Inform 2013 Jun;46(3):480-496 [FREE Full text] [doi: 10.1016/j.jbi.2013.03.008] [Medline: 23562651]
9. Jiang W, Li P, Wang S, Wu Y, Xue M, Ohno-Machado L, et al. WebGLORE: a web service for Grid LOgistic REgression. Bioinformatics 2013 Dec 15;29(24):3238-3240 [FREE Full text] [doi: 10.1093/bioinformatics/btt559] [Medline: 24072732]
10. Aono Y, Hayashi T, Phong LT, Wang L. Scalable and Secure Logistic Regression via Homomorphic Encryption. In: Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy. 2016 Presented at: DASC'16; Mar 9-11, 2016; New Orleans p. 142-144. [doi: 10.1145/2857705.2857731]
11. Aono Y, Hayashi T, Phong L, Wang L. Privacy-preserving logistic regression with distributed data sources via homomorphic encryption. IEICE Trans Inf Syst 2016 Aug;E99.D(8):2079-2089. [doi: 10.1587/transinf.2015inp0020]
12. Cheon JH, Kim D, Kim Y, Song Y. Ensemble method for privacy-preserving logistic regression based on homomorphic encryption. IEEE Access 2018 Aug;6(8):46938-46948. [doi: 10.1109/access.2018.2866697]
13. Kim M, Song Y, Wang S, Xia Y, Jiang X. Secure logistic regression based on homomorphic encryption: design and evaluation. JMIR Med Inform 2018 Apr 17;6(2):e19 [FREE Full text] [doi: 10.2196/medinform.8805] [Medline: 29666041]
14. Bonte C, Vercauteren F. Privacy-preserving logistic regression training. BMC Med Genomics 2018 Oct 11;11(Suppl 4):86 [FREE Full text] [doi: 10.1186/s12920-018-0398-y] [Medline: 30309364]
15. Chen H, Gilad-Bachrach R, Han K, Huang Z, Jalali A, Laine K, et al. Logistic regression over encrypted data from fully homomorphic encryption. BMC Med Genomics 2018 Oct 11;11(Suppl 4):81 [FREE Full text] [doi: 10.1186/s12920-018-0397-z] [Medline: 30309350]

16. Kim A, Song Y, Kim M, Lee K, Cheon JH. Logistic regression model training based on the approximate homomorphic encryption. BMC Med Genomics 2018 Oct 11;11(Suppl 4):83 [FREE Full text] [doi: 10.1186/s12920-018-0401-7] [Medline: 30309349]
17. Han K, Hong S, Cheon JH, Park D. Logistic Regression on Homomorphic Encrypted Data at Scale. In: Proceedings of the Thirty-First Innovative Applications of Artificial Intelligence Conference. 2019 Presented at: AAIC'19; January 28–30, 2019; Honolulu p. 9466-9471. [doi: 10.1609/aaai.v33i01.33019466]
18. El Emam K, Samet S, Arbuckle L, Tamblyn R, Earle C, Kantarcioglu M. A secure distributed logistic regression protocol for the detection of rare adverse drug events. J Am Med Inform Assoc 2013 May 1;20(3):453-461 [FREE Full text] [doi: 10.1136/amiajnl-2011-000735] [Medline: 22871397]
19. Jiang Y, Hamer J, Wang C, Jiang X, Kim M, Song Y, et al. SecureLR: secure logistic regression model via a hybrid cryptographic protocol. IEEE/ACM Trans Comput Biol Bioinform 2019;16(1):113-123. [doi: 10.1109/TCBB.2018.2833463] [Medline: 29994005]
20. Chen G, Chen S, Xiao Y, Zhang Y, Lin Z, Lai T. SgXpectre: stealing intel secrets from SGX enclaves via speculative execution. IEEE Secur Privacy 2020 May;18(3):28-37. [doi: 10.1109/msec.2019.2963021]
21. López-Alt A, Tromer E, Vaikuntanathan V. On-the-fly Multiparty Computation on the Cloud via Multikey Fully Homomorphic Encryption. In: Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing. 2012 Presented at: ACM'12; May 11-14, 2012; New York, USA p. 1219-1234. [doi: 10.1145/2213977.2214086]
22. Mouchet C, Troncoso-Pastoriza J, Hubaux JP. Computing across trust boundaries using distributed homomorphic cryptography. IACR Cryptol ePrint Arch 2019;2019:961 [FREE Full text]
23. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]
24. Paverd AJ, Martin A, Brown I. Modelling and Automatically Analysing Privacy Properties for Honest-but-Curious Adversaries. University of Oxford. URL: https://www.cs.ox.ac.uk/people/andrew.paverd/casper/casper-privacy-report.pdf [accessed 2020-06-17]
25. Fan J, Vercauteren F. Somewhat Practical Fully Homomorphic Encryption. IACR Cryptol ePrint Arch 2012;2012:144 [FREE Full text]
26. Brakerski Z. Fully Homomorphic Encryption without Modulus Switching from Classical GapSVP. In: Advances in Cryptology. 2012 Presented at: CRYPTO'12; August 19-23, 2012; Santa Barbara p. 868-886. [doi: 10.1007/978-3-642-32009-5_50]
27. Paillier P. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In: Advances in Cryptology. 1999 Presented at: EUROCRYPT'99; May 2–6, 1999; Prague p. 223-238. [doi: 10.1007/3-540-48910-x_16]
28. Smart NP, Vercauteren F. Fully homomorphic SIMD operations. Des Codes Cryptogr 2012 Jul 4;71(1):57-81. [doi: 10.1007/s10623-012-9720-4]
29. Bajard JC, Eynard J, Hasan MA, Zucca V. A Full RNS Variant of FV Like Somewhat Homomorphic Encryption Schemes. In: Selected Areas in Cryptography. 2016 Presented at: SAC'16; August 10-12, 2016; St. John's p. 423-442. [doi: 10.1007/978-3-319-69453-5_23]
30. Armknecht F, Boyd C, Carr C, Gjøsteen K, Jaschke A, Reuter CA, et al. A guide to fully homomorphic encryption. IACR Cryptol ePrint Arch 2015;2015:1192 [FREE Full text]
31. Ypma TJ. Historical development of the Newton–Raphson method. SIAM Rev 1995 Dec;37(4):531-551. [doi: 10.1137/1037125]
32. Tian Y, Shang Y, Tong D, Chi S, Li J, Kong X, et al. POPCORN: a web service for individual PrognOsis prediction based on multi-center clinical data CollabORatioN without patient-level data sharing. J Biomed Inform 2018 Oct;86:1-14 [FREE Full text] [doi: 10.1016/j.jbi.2018.08.008] [Medline: 30103028]
33. Breast Cancer. UCI Machine Learning Repository: Data Sets. URL: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer [accessed 2020-06-17]
34. Albrecht M, Chase M, Chen H, Ding J, Goldwasser S, Gorbunov S, et al. Homomorphic Encryption Standard. Homomorphic Encryption Standardization. URL: http://homomorphicencryption.org/wp-content/uploads/2018/08/HomomorphicEncryptionStandard2018-08-30.pdf [accessed 2020-06-17]
35. Lu Y. LR-THE. GitHub. URL: https://github.com/luyao2211/LR-THE [accessed 2020-09-12]
36. Laine K. Simple Encrypted Arithmetic Library 2.3.1. Microsoft. URL: https://www.microsoft.com/en-us/research/uploads/prod/2017/11/sealmanual-2-3-1.pdf [accessed 2020-06-17]
37. Goldreich O. Foundations of Cryptography. New York, USA: Cambridge University Press; 2004.
38. Cheon JH, Kim A, Kim M, Song Y. Homomorphic Encryption for Arithmetic of Approximate Numbers. In: Advances in Cryptology. 2017 Presented at: ASIACRYPT'17; December 3-7, 2017; Hong Kong p. 409-437. [doi: 10.1007/978-3-319-70694-8_15]
39. Lindell Y. How to simulate it – a tutorial on the simulation proof technique. In: Tutorials on the Foundations of Cryptography. Switzerland: Springer, Cham; Apr 6, 2017:277-346.

Abbreviations

BFV: Brakerski/Fan-Vercauteren
CKKS: Cheon-Kim-Kim-Song
CRT: Chinese remainder theorem
FN: false negative
FP: false positive
NTT: number theoretic transform
RLWE: ring learning with errors
SF: scaling factor
SGX: software guard extensions
SIMD: single instruction, multiple data
TN: true negative
TP: true positive

Edited by K El Emam; submitted 16.07.20; peer-reviewed by CJ Wu, A Essex, P Thaine; comments to author 10.08.20; revised version received 02.10.20; accepted 06.11.20; published 08.12.20.

Please cite as:
Lu Y, Zhou T, Tian Y, Zhu S, Li J
Web-Based Privacy-Preserving Multicenter Medical Data Analysis Tools Via Threshold Homomorphic Encryption: Design and Development Study
J Med Internet Res 2020;22(12):e22555
URL: http://www.jmir.org/2020/12/e22555/
doi:10.2196/22555
PMID:33289676

Original Paper

De-Identification of Facial Features in Magnetic Resonance Images: Software Development Using Deep Learning Technology

Yeon Uk Jeong[1], BSc, PharmD; Soyoung Yoo[2], PhD; Young-Hak Kim[3,4*], MD, PhD; Woo Hyun Shim[1,5*], PhD

[1]Department of Medical Science, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[2]Human Research Protection Center, Asan Institute of Life Sciences, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[3]Division of Cardiology, Department of Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[4]Department of Information Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[5]Department of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Woo Hyun Shim, PhD
Department of Medical Science
Asan Medical Institute of Convergence Science and Technology, Asan Medical Center
University of Ulsan College of Medicine
88, Olympic-ro 43-Gil, Songpa-gu
Seoul, 05505
Republic of Korea
Phone: 82 2 3010 2775
Email: swh@amc.seoul.kr

Abstract

Background: High-resolution medical images that include facial regions can be used to recognize the subject's face when reconstructing 3-dimensional (3D)-rendered images from 2-dimensional (2D) sequential images, which might constitute a risk of infringement of personal information when sharing data. According to the Health Insurance Portability and Accountability Act (HIPAA) privacy rules, full-face photographic images and any comparable image are direct identifiers and considered as protected health information. Moreover, the General Data Protection Regulation (GDPR) categorizes facial images as biometric data and stipulates that special restrictions should be placed on the processing of biometric data.

Objective: This study aimed to develop software that can remove the header information from Digital Imaging and Communications in Medicine (DICOM) format files and facial features (eyes, nose, and ears) at the 2D sliced-image level to anonymize personal information in medical images.

Methods: A total of 240 cranial magnetic resonance (MR) images were used to train the deep learning model (144, 48, and 48 for the training, validation, and test sets, respectively, from the Alzheimer's Disease Neuroimaging Initiative [ADNI] database). To overcome the small sample size problem, we used a data augmentation technique to create 576 images per epoch. We used attention-gated U-net for the basic structure of our deep learning model. To validate the performance of the software, we adapted an external test set comprising 100 cranial MR images from the Open Access Series of Imaging Studies (OASIS) database.

Results: The facial features (eyes, nose, and ears) were successfully detected and anonymized in both test sets (48 from ADNI and 100 from OASIS). Each result was manually validated in both the 2D image plane and the 3D-rendered images. Furthermore, the ADNI test set was verified using Microsoft Azure's face recognition artificial intelligence service. By adding a user interface, we developed and distributed (via GitHub) software named "Deface program" for medical images as an open-source project.

Conclusions: We developed deep learning–based software for the anonymization of MR images that distorts the eyes, nose, and ears to prevent facial identification of the subject in reconstructed 3D images. It could be used to share medical big data for secondary research while making both data providers and recipients compliant with the relevant privacy regulations.

(J Med Internet Res 2020;22(12):e22739) doi:10.2196/22739

KEYWORDS

de-identification; privacy protection; personal information protection; medical image; deep learning; facial feature detection; HIPAA; GDPR

Introduction

It is becoming important to handle and share big data in the health care field, and accordingly, there is a big trend to share and protect individual patient data for secondary research [1-3]. To utilize big data, data anonymization is necessary so as not to violate laws for personal privacy such as those stipulated by the Health Insurance Portability and Accountability Act (HIPAA) in the United States and General Data Protection Regulation (GDPR) in Europe [4,5]. There is a trade-off between data usability and privacy protection. Nevertheless, sufficient administrative and technical measures for previously collected information in accordance with personal information protection regulations are necessary when using the information secondarily without consent.

High-resolution magnetic resonance (MR) images of the head risk exposing a subject's face, which can be regarded at the level of photography by facial reconstruction [6]. According to HIPAA's privacy rules, full-face photographic images and any comparable images are considered to be protected health information (Multimedia Appendix 1). Budin et al [7] tested human observer recognition of 3-dimensional (3D)-rendered MR images and reported that the likelihood of correctly matching a 3D-rendered face image with a portrait of that person is higher than random guessing. Additionally, anyone can reproduce the 3D facial image from head MR images through 3D volume rendering using freeware. Therefore, it is necessary to anonymize medical images that include the face.

Facial image anonymization is not fully conducted in public medical image repositories, while some public databases even provide the original images. For example, the Alzheimer's Disease Neuroimaging Initiative (ADNI) [8] and Open Access Series of Imaging Studies (OASIS) [9] usually anonymize only metadata, while the original MR images are shared in a nonanonymized form. Anonymizing only the metadata from the medical image is not sufficient to prevent identification from the remaining medical images after removing the metadata, and existing anonymizing software is rarely used to prevent the possibility of recognition due to concerns over the deterioration of the brain image quality [10].

Previous approaches to anonymizing faces in medical images usually remove the entire facial region using a voxel classifier and mask the brain to preserve the brain image using a skull stripping technique or a convex hull [11,12]. However, since using a voxel classifier and skull stripping can be affected by variation in the characteristics of the MR images, they can produce unexpected results from heterogeneous MR image data [13]. In addition, the solution of cutting off the face has the limitation of information loss concerning the eye orbits, nasal cavity, and other underlying structures [6]. In anonymization work for medical image sharing, consistent processing of heterogeneous data and minimizing data loss will help researchers using secondary data.

The aim of this study was to develop software that can selectively distort the eyes, nose, and ears, which are the main factors for identifying a face, and make a robust anonymization algorithm that can be used on various MR images.

Methods

Defacing Process Overview

Figure 1 schematically illustrates our Deface program development process (Figure 1A) and an application example (Figure 1B). We created a deep learning model that learns the labels of the eyes, nose, and ears. The training set consisted of 3D cranial MR images and manually marked regions corresponding to each MR image. We implemented data augmentation to increase the diversity of the training data. The deep learning model was developed based on a 3D convolutional neural network. The trained model, called a "facial feature detector," can detect the eyes, nose, and ears in a 3D MR image. After the regions of the facial features have been obtained from a nonanonymous 3D MR image through the facial feature detector, the regions are anonymized according to each characteristic.

Figure 1. Process of (A) developing the facial feature detector, which is a deep learning model that can detect the eyes, nose, and ears in 3-dimensional (3D) magnetic resonance (MR) images, and (B) distorting the facial features in nonanonymized cranial MR images.

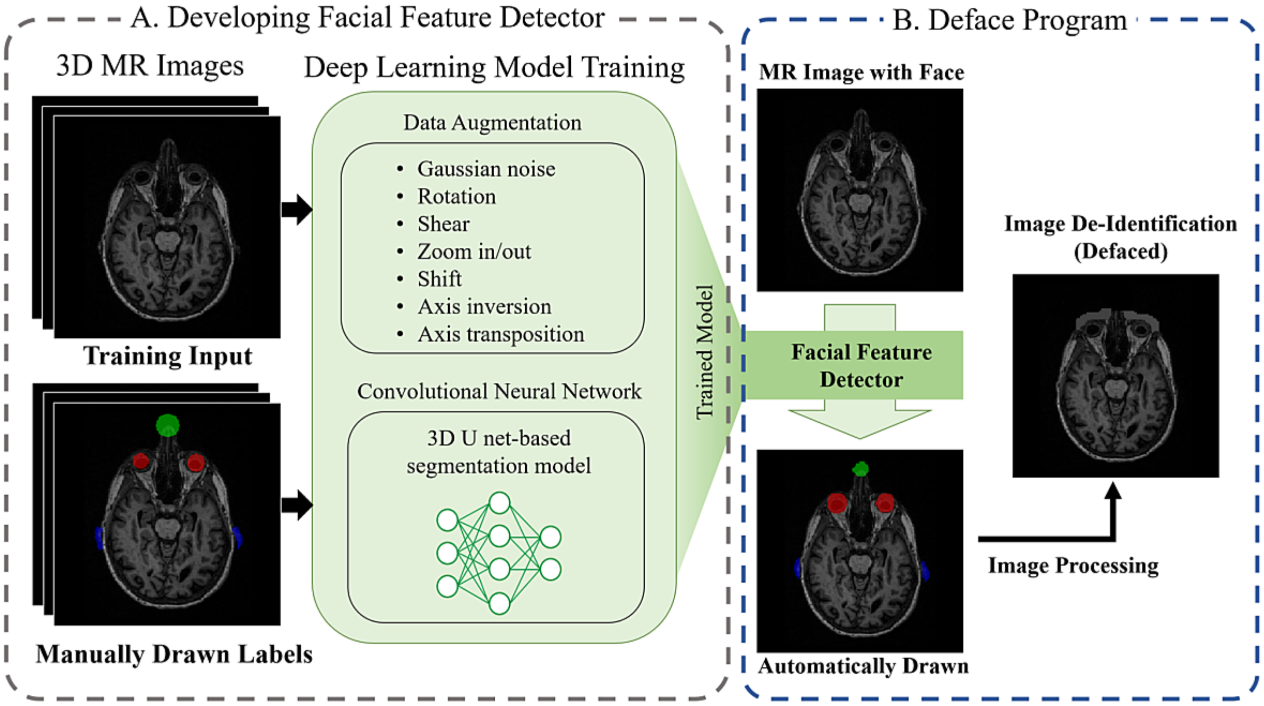


Image Acquisition

The Neuroimaging Informatics Technology Initiative (NIFTI) and Digital Imaging and Communications in Medicine (DICOM) formats of MR imaging (MRI) files were collected from the ADNI database (Magnetization Prepared RApid Gradient Echo [MPRAGE] scans; voxel size: 1.0 x 1.0 x 1.2 mm; inplane resolution: 1.0 x 1.0 mm^2; interslice spacing: 1.2 mm; field of view [FOV]: 240 x 256 x 160 mm). A total of 240 NIFTI format files were used in the creation of the deep learning model: 144, 48, and 48 for the training, validation, and test sets, respectively.

Other NIFTI-format MRI files were collected from the OASIS-3 database for use as the external test set. The 100 MR images differed in orientation, resolution, and intensity from those in the ADNI data (MPRAGE scans; voxel size: 1.0 x 1.0 x 1.0 mm; FOV: 176 x 256 x 256 mm).

Labeling

In general, supervised learning requires pairs consisting of the input object and the desired output value. In this study, the input object is a 3D cranial MR image, and the output values are regions containing the eyes, nose, or ears (the facial features). We manually drew labels that were the same as the desired output values in all of the ADNI and 20 OASIS-3 images using the AFNI program [14]. In Figure 1A, the manually drawn labels show the eyes (red) and nose (green), which are marked as spherical shapes at the corresponding positions, and the ears (blue), which are marked as the auricle regions. Each center point of the eyes and nose area was labeled in the form of a sphere. Since ears have different sizes and shapes for each person, only the auricle of the ear was segmented and labeled.

Data Augmentation

Three image augmentations were performed per 1 image in the training set. The augmented images were randomly transformed and then used for model training. As a result, 576 images per epoch were trained. Data augmentation was performed by filtering Gaussian noise, rotating from –15° to +15° around each axis in the image, randomly flipping each axis, randomly transposing between the axes, shifting each axis from 0 to 0.10, shearing each axis from 0 to 0.20, and resizing the image from 0.90 to 1.10 times the original size. After executing 1 image augmentation per original image, the validation set was validated for a total of 96 images per epoch.

The Deep Learning Algorithm

The deep learning model was trained with the manually labeled data. We created a deep learning model that can generate labels similar to manually drawn labels on the regions of the eyes, nose, and ears from cranial MR image input. The basic structure of our deep learning model is attention-gated U-net [15]. The detailed structure of our model can be found in Multimedia Appendix 2.

Metric and Loss Function

In machine learning, the “loss” or “error” variable is set to achieve the goals through the training of the model. In addition, the “metric” variable indicates how much we have achieved the goals through the model training. A machine learning model has metrics to indicate the achievement rate and is trained to reduce loss.

In this study, the metric to determine whether the model can make labels similar to the manually drawn labels is the Dice coefficient, which is double the area of overlap divided by the total number of pixels or voxels in both images: It returns 1 if the predicted regions of the model exactly match the correct answers from the labels and 0 if the regions do not overlap. When the region of the label is *Y* and the region predicted by the trained model is *X*, the Dice coefficient can be represented by:

This can also be expressed as:

where *TP* is the number of true positives, *FP* is the number of false positives, and *FN* is the number of false negatives.

The loss function in our model was: 1 – the Dice coefficient + 0.1 × categorical cross-entropy. Categorical cross-entropy is the loss function mainly used in multiclass classification, and it induces our model to learn to distinguish whether a specific pixel is from the eye, nose, ear, or another area. This model computes the loss function between the correct answer labels and the predictive labels and is trained in the direction of loss reduction (toward zero).

The model calculated Dice coefficients for 96 images in the validation set for each epoch. After 5 epochs at the highest metric score, learning was stopped when there was no further improvement.

Image Processing

Here, we describe the process of image anonymization based on the output of the facial feature detector. The deep learning model was trained by identifying the eyes, nose, and ears (5 regions), after which the program proceeded with the image anonymization process.

Identification of the eyes, nose, and ears was automatically conducted on different images according to each feature by the deep learning algorithm. The detection region for the eyes is a spherical area covering the eyeball and the skin around the eye. The process of anonymizing the surface of the eye consists of 2 steps. First, based on the detection regions for the eyes, 2 boxes capable of covering the periocular area (the skin around the eyes) are formed. Second, the contour of the face surface was obtained within the range of the boxes, and a range of ±2 voxels along each axis from that surface was modified to the same intensity value. The nose was processed by removing the image and setting the intensity of the voxels to 0 in the area where the binding box for the detected region was doubled to each side. The detection region for the ears is the protruding part called the auricle. For the anonymization of the ears, random values were assigned to each voxel of the detection regions of the ears, and those values were generated in the noise range of the air in the MR image.

In the case of the medical images in DICOM format, it is necessary to anonymize the personal information in the header, and so we carried this out on the 20 DICOM headers using the Deface program (the DICOM headers are listed in Multimedia Appendix 3). The list was selected based on the HIPAA safe harbor provision [16].

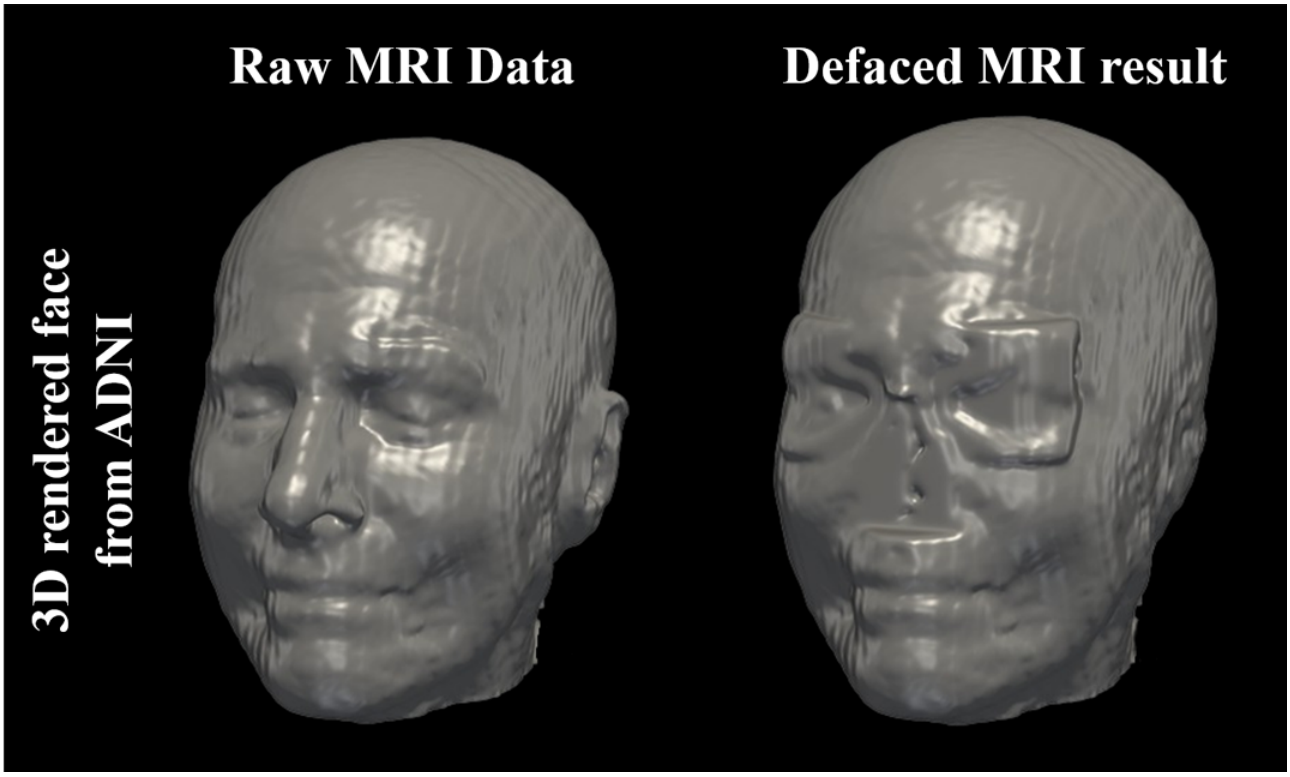
Results

In the 23rd epoch, the average Dice coefficient of the validation set was the highest at 0.821. In the 28th epoch, the training of the model was stopped because the Dice score of the validation set did not improve. The average Dice score of 576 images trained over 23 epochs was 0.801. The average Dice scores using the test sets comprising 48 ADNI and 20 OASIS-3 images were 0.859 and 0.794, respectively. The Deface program was applied to the ADNI data, but anonymization was performed on the OASIS-3 data without any additional manipulation.

Figure 1B shows the process of distorting a sample nonanonymized cranial MR image. Three axial views of the cross-sectional MR image were obtained from a representative image in the ADNI test set. The first is the nonanonymized cranial MR image, the second is an MR image with the detection regions (the labels of the eyes, nose, and ears predicted by the facial feature detector) as output for the facial feature detector, and the third is the final anonymized image based on the detection regions (red marks denoting the eyes, green marks denoting the nose, and blue marks denoting the ears). It took 177.91 seconds to save the detection region pictures and distorted MR images as NIFTI format files from 48 images of the ADNI test set. The image was distorted according to the characteristics of each facial feature. The 3D box space containing the entire volume of the nose was removed. The eyes were covered with similar brightness intensity on the surface. For the ears, the detection regions were replaced by space with noise.

We applied the Deface program to 48 ADNI images and 100 OASIS-3 images as the test sets and then confirmed the accuracy of distorting the facial features in the 3D reconstructions of the face (Figure 2 shows the 3D volume-rendered images). Since face reconstruction is in violation of the OASIS data use terms, OASIS data were not included in the figure. A sample image was selected from the ADNI test sets, and we compared the before and after anonymization. As shown in Figure 2, the facial features clearly identifiable in the 3D images beforehand are distorted after processing: The auricle and nose have disappeared, and the eyes appear blurry.

Figure 2. 3-dimensional (3D) volume rendering of magnetic resonance images (MRI), showing the raw and distorted images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) .



The Deface program was used to validate the de-identification performance by Microsoft Azure's facial recognition artificial intelligence service (Face detection_01 model) [17]. We found that all 48 reconstructed face images from the ADNI test set were de-identified. Although 46 unmodified images were recognized as faces and location information of face landmarks was derived, the faces in all 48 defaced images were not recognized. The other 2 unmodified images failed the face recognition process because they were noisy or parts of the face were cropped. The result of the face detection service for 1 representative image of the ADNI test set can be found in Multimedia Appendix 4.

Discussion

Principal Findings

In this study, we developed a program that can recognize the eyes, nose, and ears in MR images by applying artificial intelligence, after which they were blurred. We implemented the facial feature detector based on the 3D U-net deep learning model to automatically detect the eyes, nose, and ears. The reason for the development of this anonymization program is that 3D facial reconstruction of high-resolution MRI can show an individual's similarity to a facial photograph [6,7], which contravenes the rules for protecting personal information required by regulating bodies such as HIPAA. Anonymization is required for the sharing of medical image data so as not to infringe on the personal information rules. However, distorting images is disadvantageous for secondary research due to the loss of information, but too little distorting leads to the possibility of recognition [10]. We attempted to anonymize the face while minimizing the loss of information by modifying only the surface of the eyes, nose, and ears in cranial MR images. In addition, in the case of the DICOM format, a function to remove text including personal information that can be obtained from the header was added. We released the source code to GitHub [18].

Figure 1A shows the process of developing the facial feature detector. The cranial MR images and manually marked facial features (eyes, nose, and ears) were used as the training set. We drew different labels for each facial feature for the manually drawn labels. Although the eyes and nose can be specified in a range of only the central location information, the shape of the ear varies relatively widely among people. Furthermore, because the ears are adjacent to the brain, images of the brain can be obscured during the image distortion Therefore, only the segmented regions of the auricle were used as labels so that the program did not select regions other than the ear.

Although the training data for the deep learning model comprised 144 images from ADNI, we introduced data augmentation to achieve robust performance in other MRI standards (Figure 1A). The training set was augmented via various techniques so that the facial feature detector could show robust performance even with unknown data. We evaluated OASIS-3 data in which the adjustment, orientation, FOV, resolution, and intensity histograms were completely different from the ADNI data in the training set. We confirmed that the facial features were distorted in 100 OASIS-3 images by the MRI viewer. Labels were manually drawn on 20 OASIS-3 images, and our facial feature detector worked well, with an average Dice coefficient of 0.794. This has the potential to assist in the construction of anonymous big data with different MRI standards collected from multiple institutions.

We applied different processes to blur each facial feature location. The eyes are close to the frontal lobe, so they were distorted only along the surface. The intensity of the pixels was converted to a value similar to the surface of the skin to make it appear on the surface when 3D rendering. Since the nose is usually the most protuberant part of the face, the area that covers the entire range of the nose was deleted to make it impossible to infer the original shape of the nose. The 3D box space containing the entire volume of the nose was removed to prevent recognition via the nose shape. The ears were only segmented by the facial feature detector, so only the corresponding regions were distorted to preserve the brain image. If regions such as the shape of the ears are simply removed, the shape of the ears may be revealed by the noise from air in the MR image. We reduced the possibility of recognition by replacing the ear regions with generated random values within the air noise range of the input MR image.

3D facial reconstruction of high-resolution MRI can be generated by a freeware MRI viewer [19]. Moreover, the faces of patients in MR images from publicly available data can be revealed (Figure 2). As the OASIS-3 images are smoother than the ADNI images, they can be reconstructed with a clearer face image in the case of high-resolution MRI. However, we showed that the face could be distorted in the 3D-rendered image after applying our Deface program. Since the image was preserved except for the user-designated facial features, researchers can obtain the necessary information from MRI images without revealing the patient's identity.

Comparison With Prior Work

Previous studies have applied techniques to remove the entire face regions, and the evaluation of anonymization was via direct human observation of face landmarks [7,11,12]. In another technique, the Human Connectome Project [20], a public repository of MRI images, conducted distorting by modifying a certain thickness of the facial surface [13]. We distorted the ears in addition to the face surface, with options to blur the eyes, nose, and ears separately, as may be required when conducting secondary research. The images with eyes, nose, and ears anonymized were verified by applying a face recognition tool. Furthermore, while previous studies have applied algorithms to process single MRI datasets, our Deface program was tested on 2 different MRI datasets to improve compatibility.

Limitations

Among the facial features, wrinkles or the mouth can be identifiers but were not considered in this study. To train the deep learning model, we needed to manually draw labels that mark facial features. We are planning to construct a training dataset that takes into account additional facial features for further study. Once labeled training data comprising any desired facial feature have been constructed, our facial feature detector can evolve through deep learning.

Conclusions

Patients' faces can be reconstructed from high-resolution cranial MR images at the photograph level, so there is a risk of infringing the personal information rules prescribed by HIPAA and GDPR when sharing data. Hence, we suggested a method to perceive the facial features in MR images via deep learning technology to specifically blur certain facial features. Users can create anonymization regions that blur the desired parts of the patient's face (eyes, nose, or ears), which helps provide data for secondary research without violating relevant personal information regulations.

Acknowledgments

This research was supported by an Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (2018-0-00861), Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03030713), and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI18C2383). We thank the Biomedical Computing core facility at the ConveRgence mEDIcine research cenTer (CREDIT), Asan Medical Center for their technical support and instrumentation funded by Asan Institute for Life Sciences (2018-776).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Information Protection Regulations.
[DOCX File , 21 KB - jmir_v22i12e22739_app1.docx]

Multimedia Appendix 2
Deep learning model structure.
[DOCX File , 127 KB - jmir_v22i12e22739_app2.docx]

Multimedia Appendix 3
DICOM header with personal information.

[DOCX File , 14 KB - jmir_v22i12e22739_app3.docx]

Multimedia Appendix 4
Face recognition test.
[DOCX File , 249 KB - jmir_v22i12e22739_app4.docx]

References

1. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
2. Vallance P, Chalmers I. Secure use of individual patient data from clinical trials. The Lancet 2013 Sep 28;382(9898):1073-1074. [doi: 10.1016/S0140-6736(13)62001-2] [Medline: 24075034]
3. El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. BMJ 2015 Mar 20;350:h1139 [FREE Full text] [doi: 10.1136/bmj.h1139] [Medline: 25794882]
4. Chevrier R, Foufi V, Gaudet-Blavignac C, Robert A, Lovis C. Use and Understanding of Anonymization and De-Identification in the Biomedical Literature: Scoping Review. J Med Internet Res 2019 May 31;21(5):e13484 [FREE Full text] [doi: 10.2196/13484] [Medline: 31152528]
5. Kayaalp M. Patient Privacy in the Era of Big Data. Balkan Med J 2018 Jan 20;35(1):8-17 [FREE Full text] [doi: 10.4274/balkanmedj.2017.0966] [Medline: 28903886]
6. Prior FW, Brunsden B, Hildebolt C, Nolan TS, Pringle M, Vaishnavi SN, et al. Facial recognition from volume-rendered magnetic resonance imaging data. IEEE Trans Inf Technol Biomed 2009 Jan;13(1):5-9. [doi: 10.1109/TITB.2008.2003335] [Medline: 19129018]
7. Budin F, Zeng D, Ghosh A, Bullitt E. Preventing facial recognition when rendering MR images of the head in three dimensions. Med Image Anal 2008 Jun;12(3):229-239 [FREE Full text] [doi: 10.1016/j.media.2007.10.008] [Medline: 18069044]
8. Alzheimer's disease neuroimaging initiative. 2020. URL: http://adni.loni.usc.edu/ [accessed 2019-06-05]
9. OASIS: Open access series of imaging studies. 2020. URL: https://www.oasis-brains.org/ [accessed 2019-08-20]
10. Schwarz CG, Kremers WK, Therneau TM, Sharp RR, Gunter JL, Vemuri P, et al. Identification of Anonymous MRI Research Participants with Face-Recognition Software. N Engl J Med 2019 Oct 24;381(17):1684-1686 [FREE Full text] [doi: 10.1056/NEJMc1908881] [Medline: 31644852]
11. Schimke N, Kuehler M, Hale J. Preserving privacy in structural neuroimages. In: Li Y, editor. Data and Applications Security and Privacy XXV. Heidelberg, Germany: Springer Publishing Company; 2011:301-308.
12. Bischoff-Grethe A, Ozyurt IB, Busa E, Quinn BT, Fennema-Notestine C, Clark CP, et al. A technique for the deidentification of structural brain MR images. Hum Brain Mapp 2007 Sep;28(9):892-903 [FREE Full text] [doi: 10.1002/hbm.20312] [Medline: 17295313]
13. Milchenko M, Marcus D. Obscuring surface anatomy in volumetric imaging data. Neuroinformatics 2013 Jan;11(1):65-75 [FREE Full text] [doi: 10.1007/s12021-012-9160-3] [Medline: 22968671]
14. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 1996 Jun;29(3):162-173. [doi: 10.1006/cbmr.1996.0014] [Medline: 8812068]
15. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: Learning to leverage salient regions in medical images. Med Image Anal 2019 Apr;53:197-207 [FREE Full text] [doi: 10.1016/j.media.2019.01.012] [Medline: 30802813]
16. Title 45, Subtitle A, Subchapter C, Part 164, Subpart E, §164.514. Electronic code of federal regulations. 2020. URL: https://www.ecfr.gov/cgi-bin/text-idx?SID=20e0360351a51dd55ee6e80cc9aae47c&node=se45.1.164_1514&rgn=div8 [accessed 2020-11-28]
17. Facial recognition AI service. Microsoft Azure. 2020. URL: https://azure.microsoft.com/en-us/services/cognitive-services/face/#demo [accessed 2020-11-28]
18. yeonuk-Jeong / De-facer: De-identifier reconstructable facial information in Medical image (CT, MRI). GitHub. 2020. URL: https://github.com/yeonuk-Jeong/Defacer [accessed 2020-11-28]
19. Debus C, Floca R, Ingrisch M, Kompan I, Maier-Hein K, Abdollahi A, et al. MITK-ModelFit: A generic open-source framework for model fits and their exploration in medical imaging - design, implementation and application on the example of DCE-MRI. BMC Bioinformatics 2019 Jan 16;20(1):31 [FREE Full text] [doi: 10.1186/s12859-018-2588-1] [Medline: 30651067]
20. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K, WU-Minn HCP Consortium. The WU-Minn Human Connectome Project: an overview. Neuroimage 2013 Oct 15;80:62-79 [FREE Full text] [doi: 10.1016/j.neuroimage.2013.05.041] [Medline: 23684880]

Abbreviations

2D: two-dimensional

3D: three-dimensional
ADNI: Alzheimer's Disease Neuroimaging Initiative
DICOM: Digital Imaging and Communications in Medicine
FOV: field of view
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
MPRAGE: Magnetization Prepared RApid Gradient Echo
MR: magnetic resonance
MRI: magnetic resonance imaging
NIFTI: Neuroimaging Informatics Technology Initiative
OASIS: Open Access Series of Imaging Studies

Edited by G Eysenbach, R Kukafka; submitted 22.07.20; peer-reviewed by D Marcus, A Phillips, P N B; comments to author 19.08.20; revised version received 09.09.20; accepted 12.11.20; published 10.12.20.

Please cite as:
Jeong YU, Yoo S, Kim YH, Shim WH
De-Identification of Facial Features in Magnetic Resonance Images: Software Development Using Deep Learning Technology
J Med Internet Res 2020;22(12):e22739
URL: http://www.jmir.org/2020/12/e22739/
doi:10.2196/22739
PMID:33208302

Original Paper

Communicating With Patients About Software for Enhancing Privacy in Secondary Database Research Involving Record Linkage: Delphi Study

Cason Schmit[1], JD; Kobi V Ajayi[1], MPH, MBA; Alva O Ferdinand[2], JD, DrPH; Theodoros Giannouchos[1,3], PhD, MS; Gurudev Ilangovan[1], MSCS, MEng; W Benjamin Nowell[4], PhD; Hye-Chung Kum[1], PhD

[1]Population Informatics Lab, Department of Health Policy & Management, Texas A&M University School of Public Health, College Station, TX, United States

[2]Southwest Rural Health Research Center, Department of Health Policy & Management, Texas A&M University School of Public Health, College Station, TX, United States

[3]Pharmacotherapy Outcomes Research Center, College of Pharmacy, University of Utah, Salt Lake City, UT, United States

[4]Global Healthy Living Foundation, Upper Nyack, NY, United States

Corresponding Author:
Cason Schmit, JD
Population Informatics Lab
Department of Health Policy & Management
Texas A&M University School of Public Health
121 SPH Admin Building
212 Adriance Lab Road
College Station, TX, 77843
United States
Phone: 1 9794360277
Email: schmit@tamu.edu

Abstract

Background: There is substantial prior research on the perspectives of patients on the use of health information for research. Numerous communication barriers challenge transparency between researchers and data participants in secondary database research (eg, waiver of informed consent and knowledge gaps). Individual concerns and misconceptions challenge the trust in researchers among patients despite efforts to protect data. Technical software used to protect research data can further complicate the public's understanding of research. For example, MiNDFIRL (Minimum Necessary Disclosure For Interactive Record Linkage) is a prototype software that can be used to enhance the confidentiality of data sets by restricting disclosures of identifying information during the record linkage process. However, software, such as MiNDFIRL, which is used to protect data, must overcome the aforementioned communication barriers. One proposed solution is the creation of an interactive web-based frequently asked question (FAQ) template that can be adapted and used to communicate research issues to data subjects.

Objective: This study aims to improve communication with patients and transparency about how complex software, such as MiNDFIRL, is used to enhance privacy in secondary database studies to maintain the public's trust in researchers.

Methods: A Delphi technique with 3 rounds of the survey was used to develop the FAQ document to communicate privacy issues related to a generic secondary database study using the MiNDFIRL software. The Delphi panel consisted of 38 patients with chronic health conditions. We revised the FAQ between Delphi rounds and provided participants with a summary of the feedback. We adopted a conservative consensus threshold of less than 10% negative feedback per FAQ section.

Results: We developed a consensus language for 21 of the 24 FAQ sections. Participant feedback demonstrated preference differences (eg, brevity vs comprehensiveness). We adapted the final FAQ into an interactive web-based format that 94% (31/33) of the participants found helpful or very helpful. The template FAQ and MiNDFIRL source code are available on GitHub. The results indicate the following patient communication considerations: patients have diverse and varied preferences; the tone is important but challenging; and patients want information on security, identifiers, and final disposition of information.

Conclusions: The findings of this study provide insights into what research-related information is useful to patients and how researchers can communicate such information. These findings align with the current understanding of health literacy and its challenges. Communication is essential to transparency and ethical data use, yet it is exceedingly challenging. Developing FAQ template language to accompany a complex software may enable researchers to provide greater transparency when informed consent is not possible.

(J Med Internet Res 2020;22(12):e20783) doi:10.2196/20783

KEYWORDS

Delphi technique; privacy; communication barriers; medical record linkage; research subjects; big data

Introduction

Transparency and Trust in Secondary Database Research

The researcher-participant relationship rests on a fragile foundation, weakened by a history of scandal and abuse. Past research abuses might not be readily remembered by the general public, but the scars remain in the social subconscious [1-4]. Although many still view research positively, transparency between researchers and the study participants is a critical element in building and maintaining publics' trust [5]. Transparency is key to supporting informed consent and *respect for persons*, the central ethical principle of the bioethical framework that governs human subjects research [6,7].

Unfortunately, it is difficult to achieve transparency in secondary database research because ethical review bodies (ie, institutional review boards [IRBs]) frequently waive informed consent requirements. This is done because informed consent is often impractical (ie, no contact between researchers and data subjects), and it is not possible to know the purpose of data use at all instances at the time of data collection [8]. Without the traditional informed consent process, secondary database researchers have trouble cultivating strong researcher-participant relationships.

The lack of informed consent is not without consequence. Individuals want to know how their data are used and want to be partners in the research process [5,9]. Moreover, people might be reluctant to participate in research when they fear that researchers are taking advantage of their information [9]. Survey evidence suggests that the public wants researchers to do a better job at communicating with the public [9,10]. A survey of 3516 patients suggests that communication about data protection methods would help improve comfort levels of the public with research [10]. This study also found that comfort levels of the public may be improved by using methods that minimize the exposure of unique identifiers while linking data for research purposes [10]. These findings suggest that a privacy statement that increases database research transparency and discusses the software used to enhance privacy would increase publics' trust.

However, communication between researchers and the public can be challenging [9]. The expertise of researchers often obstructs effective communication. Jargon, technical language, and complex concepts are barriers to understanding. In secondary database research, these challenges are immense. Explanations of critical concepts often require additional explanations of related concepts. For example, a participant might be interested in knowing why a researcher needs a specific data element (eg, names); this might require further explanation of related issues such as record linkage, which invites additional questions and explanations of other concepts (eg, linked data) and additional tangential issues (eg, storage, maintenance, and data reuse).

Transparency in secondary database research requires communicating research risks, including privacy and confidentiality of personal data. Data subjects want to understand how researchers use and safeguard their data. This can include employing technology and software to enhance the privacy, confidentiality, and security of sensitive information. Thus, describing how technology is used to safeguard participant data will likely promote transparency and hopefully increase trust between researchers and data subjects.

For example, we have designed and evaluated a prototype user interface called MiNDFIRL (Minimum Necessary Disclosure For Interactive Record Linkage) to enhance data set confidentiality by restricting disclosures of identifying information while linking records. One study found that the on-demand interactive interface of MiNDFIRL reduced visible characters in the identifying data elements used in record linkage by 92.15% as compared with traditional manual record linkage, with little to no impact on decision quality or completion time [11,12]. In our 2 case studies with real data from 2 teaching hospitals, used in conjunction with automated record linkage software, we found that the prototype MiNDFIRL software could effectively support high-quality record linkage with much less data disclosure than manual record linkage [13]. Disclosing how a software, such as MiNDFIRL, protects research data can help cultivate publics' trust in researchers.

Transparency on these issues is imperative as research data are particularly sensitive [9]. Secondary database researchers frequently use health information because of its importance in understanding critical societal issues, including effective treatments, health care costs, and service utilization. However, health data also carries a risk of social, economic, and psychological harm. Researchers often disclose the potential benefits and harms of the study to the study participants in the informed consent process. However, this is more difficult in secondary database research because of the lack of direct contact with data subjects.

Improving Transparency in Secondary Database Research

In nonresearch settings, it is common to disclose privacy and confidentiality practices in a public privacy statement on the premise that transparency coupled with accountability is an effective form of privacy protection and ethical data use [14,15]. In theory, privacy statements provide an opportunity to increase public transparency. However, privacy statements in practice are long, technical, and burdensome to process and understand.

Frequently asked question (FAQ) documents are another mechanism used to promote transparency and understanding. In contrast to a privacy statement, which forces patients to find answers to their specific questions, FAQs present common questions and provide direct answers [16]. If a person only has a few questions they want answered, an FAQ can help them quickly find answers without reading volumes of irrelevant information. Some research institutions (eg, Texas A&M University) use a question and answer format, similar to an FAQ, in the informed consent templates for research (ie, “What are the risks of participating in this research?”).

This study aims to improve communication with patients about how a complex software, such as MiNDFIRL, is used to enhance privacy in secondary database studies to maintain publics’ trust in researchers. To do this, we aim to develop a template FAQ language that describes key issues in secondary database research to the public, including the necessity of sensitive identifiers (eg, names) in the record linkage process and the safeguards used by researchers to protect data from research risks (eg, researchers identifying their neighbor in a data set). To do this, we aim to develop a template FAQ language that describes key issues in secondary database research to the public, including the necessity of sensitive identifiers (eg, names) in the record linkage process and the safeguards used by researchers to protect data from research risks (eg, researchers identifying their neighbor in a data set). Through this template FAQ, we aim to improve research transparency, promote ethical data use, and cultivate trust between researchers and the public. This study focuses on the patient community and its perceptions of the benefits and risks of research using sensitive health data without informed consent.

We focus on health data because the public has heightened awareness of its sensitivity and harm [10,17]. However, many of the key concepts that we explore (eg, identifiability and record linkage) are broadly applicable to data obtained from nonhealth sectors (eg, education and social services) [18-21]. Our study primarily focuses on creating a template language for researchers using the MiNDFIRL software for record linkage, but we also anticipate that many FAQ items that we created will have broad applicability to secondary database research.

Methods

Design

We conducted a Delphi study using a web-based questionnaire administered through Qualtrics XM (SAP SE). The goal of the Delphi process was to create an FAQ to anticipate and answer the questions of the data subjects related to the use of their data in secondary database research. These questions include general questions and answers related to secondary database research, record linkage, and the MiNDFIRL software (ie, facilitating record linkage while enhancing confidentiality).

The Delphi technique is useful for studying communication between researchers and data subjects for several reasons. The Delphi approach is particularly suited to investigate communication strategies with patients and data subjects, where there are differences in thought, and the body of knowledge is still developing [22,23]. In addition, the Delphi process facilitates anonymous and confidential feedback from patients with diverse perspectives without the biases common to other consensus techniques such as group discussions and interviews [24,25]. Moreover, the structured feedback in a Delphi approach prevents dominant personalities from suppressing diverse inputs and perspectives [22].

During each Delphi round, we asked the patient expert panel a series of questions about a draft FAQ document. Between each round, we revised the FAQ language and content based on the feedback of the participants. In addition to written descriptions, the FAQ drafts included visual images and a short video demonstrating how the MiNDFIRL software works during the record linkage process (referred to as *patient matching* in the FAQ) to enhance user comprehension.

The FAQ drafts contained approximately 3000 words, and the accompanying survey instrument contained approximately 1500 words. Owing to the complexity and the word length of both documents, we excluded survey results if respondents completed the study in less than 10 min, to allow for the inclusion of constructive feedback. The study received ethical approval from the IRB of Texas A&M University (IRB2019-0234).

Participants

We identified the chronic patient population (operationalized as patients with frequent encounters with the health care system) as the appropriate *expert* panel for the Delphi study. Patients with chronic conditions are the *experts* because they are likely to have conditions that are of interest to secondary database researchers, and patient voice is essential to health care, supporting systems (eg, health research), and related policy decisions [26]. In addition, patients with chronic conditions are likely to have multiple health care providers and thus require record linkage for a comprehensive understanding of their patient experience.

Recruitment

We recruited patients via email using purposive sampling from 3 patient-powered research networks (PPRNs) and employees and staff of a large university in the south. Specifically, ArthritisPower, Chronic Obstructive Pulmonary Disease PPRN, and Interactive Autism Network were PPRNs that were part of the national Patient-Centered Clinical Research Network that we collaborated with. We included participants with at least one diagnosed chronic disease and with at least two physician visits for their condition in the previous 12 months. We compensated participants on a graduated basis: US $20 for completing round 1, US $30 for completing round 2, and US $50 for completing round 3, amounting to a maximum of US $100 in total as gift cards. We invited 45 participants to join this study.

FAQ Development

Following recommendations from several health care organizations to enhance the readability level of the education material, the first FAQ had a Flesch-Kincaid readability score of 7.82 [27,28]. Findings from 3 prior studies informed the content and language of the FAQ template, including 2 nominal group technique (NGT) studies and a separate Delphi study [13,29]. One NGT study and a Delphi study used panels of experts in the ethical, legal, and social implications fields to help identify legal and ethical issues associated with secondary database research involving record linkage. Another NGT study used a panel of patients with chronic conditions to help identify issues that were important to the patient population.

Delphi Rounds Procedure

Each Delphi round contained a mix of open-ended questions and 5-point Likert scale questions (eg, strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree) administered on the web via Qualtrics. All 3 surveys were pilot tested by members of the research team before being administered. The round 1 survey included 50 questions, with fewer questions in successive rounds (Delphi survey instruments and materials are detailed in Multimedia Appendices 1-6). Participants were asked to provide feedback on the content of the FAQ document, including whether FAQ sections provided information that was understandable or important to patients. In addition, in round 1, we solicited open-ended responses after each FAQ section to identify gaps, solicit suggestions for new FAQ content (ie, “Questions 1-4 were about information and how it is used in research. Is there anything more that you would like to know on this topic?”), identify specific concerns (ie, “Do you have any concerns that you would like the FAQ to address”), and seek any other general comments. In rounds 2 and 3, the survey focused on the language and content of the FAQ that participants found problematic or where participant feedback suggested divergent opinions. In rounds 2 and 3, we also provided participants with a summary of the participant feedback from the previous round and a *tracked-changes* (eg, *redline*) version of the revised FAQs to enable them to see the specific revisions that were made based on the feedback. Participants were given just over a week to complete each round of the survey, with 2 reminder emails sent out within that time to reduce attrition.

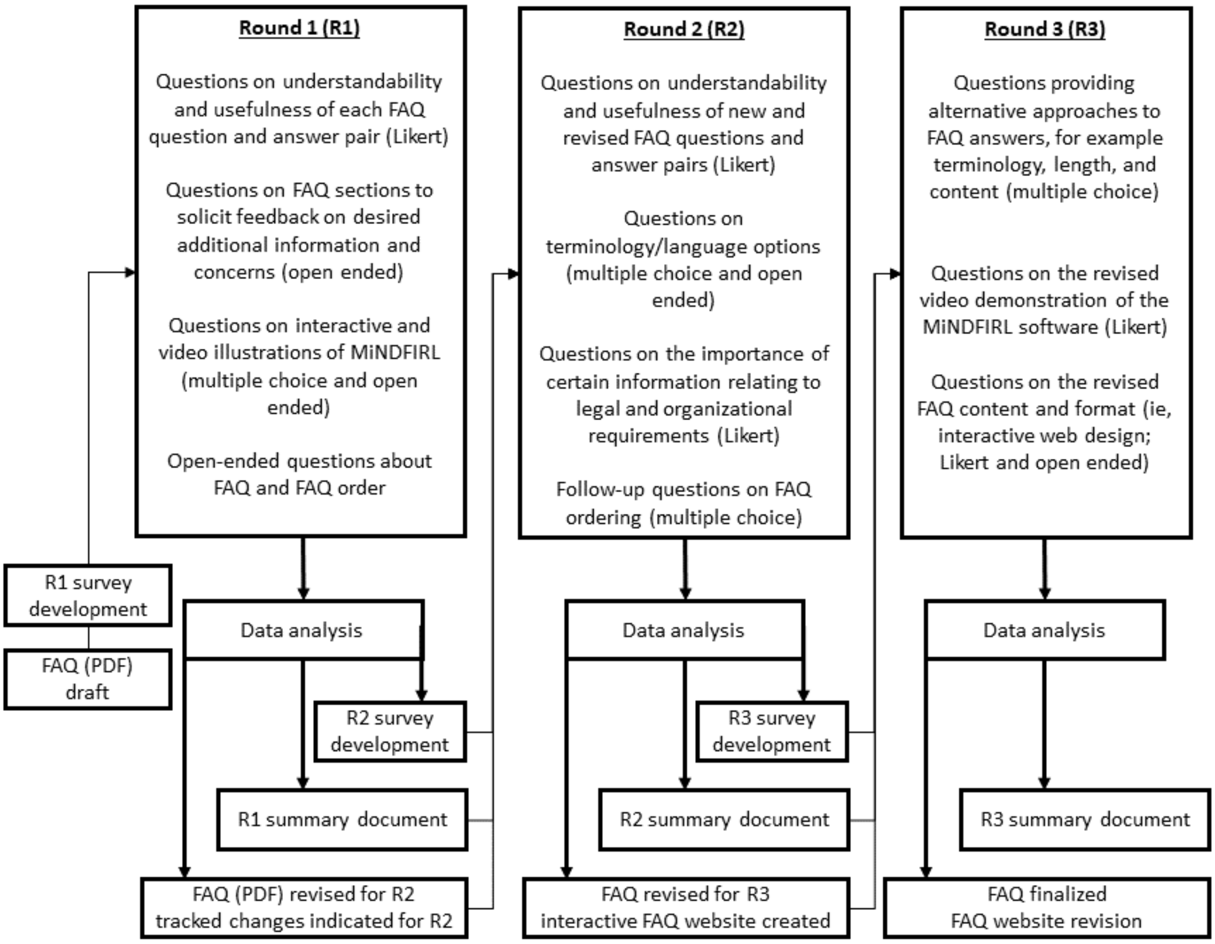
Round Analysis, FAQ Draft Revisions, and Consensus Criteria

After each Delphi round, we triaged the FAQ content as *problematic or nonproblematic*. If FAQ sections had fewer than 3 participants providing negative feedback (ie, *disagree* or *strongly disagree*), they were deemed *nonproblematic*. We viewed nonproblematic FAQ sections as a consensus language and made only minor edits to these sections (eg, revisions to terminology). FAQ sections that received negative feedback from 3 or more participants (on any question relating to understandability or importance) were deemed problematic. For all FAQ items, we examined open-ended feedback and comments to identify points of confusion (eg, terminology, unexplained concepts, and poor phrasing), requested clarifications (eg, examples), and other participant recommendations. One researcher incorporated the feedback in a redlined document alongside relevant participant comments to ensure that the revisions addressed the feedback. This document was shared with the research team for review and approval. If we made substantial revisions to an FAQ section (ie, for FAQ items deemed to be problematic), we solicited additional feedback in subsequent rounds.

In addition to survey questions about specific FAQ question and answer content, rounds 2 and 3 contained questions to identify preferred terminology and strategies for communicating key concepts to the patient community. Whenever participant feedback suggested a divergence of opinions or suggestions, questions were devised to explore the divergence by explicitly raising the issues and providing participants with the opportunity to provide additional feedback (eg, providing alternative approaches to answering a question in the FAQ sections). An overview of our process is shown in Figure 1. The survey instruments and the round summaries provided to the participants are included in Multimedia Appendices 1-6.

We note that our consensus criterion, fewer than 3 individuals providing negative feedback, is highly conservative among Delphi studies. The Delphi technique does not demand a specific threshold for consensus [22,25,30]. Some Delphi studies use simple majorities to define consensus [25,31]. In contrast, our consensus criterion requires between 90% and 92% positive or neutral feedback, depending on the round, notably higher than other Delphi studies [25].

Figure 1. Overview of the Delphi process and content for round 1 (n=38), round 2 (n=37), and round 3 (n=33). FAQ: frequently asked question; MiNDFIRL: Minimum Necessary Disclosure For Interactive Record Linkage.



Thematic Analysis

After all 3 rounds, 2 researchers conducted a preliminary inductive thematic analysis of the open-ended responses to provide an additional context for our primary result: the FAQ template document. The thematic analysis was intended to identify broader lessons learned about communicating with patients about research from the Delphi process, including common concerns, desired content, and preferred communication approaches. One researcher reviewed the open-ended comments and generated an initial list of themes. The second researcher reviewed the initial list, all participant feedback, and made revisions to the list of themes. These draft themes were reviewed and discussed in a meeting with a third researcher providing input. We then identified participant quotes that were representative of the themes.

Results

Participant Demographics

A total of 45 participants were invited to participate in the Delphi study. Of those, 38 participated in round 1, resulting in a response rate of 84% (Table 1). Females (28/38, 74%) and non-Hispanic Whites (33/38, 87%) were disproportionately represented. The age range of the participants was 21 to 78 years, with a median of 49 years. The majority had a college degree or higher education (35/38, 92%), reflecting that the patients engaged in research with PPRNs and the staff at a university that we recruited from. Less than half of the sample (15/38, 39%) had 2 to 5 physician visits, with the rest having more than 5 visits in the last year and 63% (24/38) reported having good or very good health status. The most common clinical conditions included type II diabetes, thyroid disease, rheumatoid arthritis, and chronic pain.

Table 1. Demographic information of participants who completed round 1 (n=38).

Characteristics	Values
Age (years)	
Mean (SD)	50 (14.6)
Range	21-78
Gender, n (%)	
Male	10 (26)
Female	28 (74)
Education, n (%)	
Some college credit and no degree	3 (8)
Associate degree	7 (18)
Bachelor’s degree	11 (29)
Master’s degree	11 (29)
Doctoral degree	6 (16)
Race and ethnicity, n (%)	
White	33 (87)
African American or Black	3 (8)
Asian	1 (3)
Other	1 (3)
Average physician visits in 12 months, n (%)	
2-5 times	15 (39)
6-10 times	14 (37)
>10 times	9 (24)
Self-reported health status, n (%)	
Excellent	2 (5)
Very good	10 (26)
Good	14 (37)
Fair	12 (32)

Principal Results

The principal result of this study is the final FAQ template document. The final FAQ template had a Flesch-Kincaid readability score of 8.66, which is slightly higher than the initial FAQ score of 7.82. The FAQ template is included in Multimedia Appendix 7, and it is also published on the web and accompanies the open-source software package for MiNDFIRL available on GitHub [32,33]. To the best of our knowledge, this is the first open-source software to include template documentation to facilitate transparent communication with subjects.

Delphi Round 1

In round 1 (n=38), participants responded favorably to the overall FAQ, with 89.86% (2322/2584) of responses within the strongly agree and agree category across all survey questions on the FAQ content. Most participants strongly agreed or agreed that the questions included in the FAQ were easy to understand (755/798, 94.6%), the answers provided in the FAQ were easy to understand (407/494, 82.4%), and the answers contained useful information (432/494, 87.4%).

Participants generally found the visual content to be helpful (25/38, 65% of the participants indicated that the record linkage demonstration video was helpful; 31/38, 81% of the participants indicated that the visual image showing how the MiNDFIRL software reduces the privacy risk was helpful). Negative feedback about the demonstration video and image, however, was mainly related to technical issues (eg, could not view) and included a request for more information. The question that received the strongest negative feedback (18 participants either disagreed or strongly disagreed that the FAQ item contains useful information) was “What will you do if you discover that my data has been misused?” Many participants objected to the lack of specificity with the provided answer (“While we take great measures to safeguard your data, if a data breach were to occur, we would follow legal guidelines for breach notification.”) One participant noted:

> *What are the legal guidelines for breach notification? Are there any ramifications for the institution for a breach? Are there any remediation efforts that will be undertaken?*

Of the 50 questions in round 1, 7 FAQ question and answer items were designated as *problematic* and triaged for substantial edits for round 2. In round 1, 5 participants (5/38, 13%) had concerns about the terminology used to describe information subsets (ie, *identifiers* and *nonidentifiers*).

Delphi Round 2

A total of 37 (37/38, 97%) participants were included in round 2. Round 2 had fewer questions (25 items) than round 1 because of a high level of agreement and positive feedback. In general, the revisions made before round 2 were related to terminology and improving the readability of the FAQ. In addition, round 1 responses indicated that some concepts should be explained in more detail, and therefore, several new FAQ question and answer pairs were added to enhance the understanding (eg, “What is Patient Matching?”). We also included new visuals to aid understanding of key concepts.

In round 2, we sought specific feedback for the newly created FAQ items and the 7 FAQ items that were deemed problematic in round 1. Of these, strong majorities of participants agreed that the new and revised questions were easy to understand (mean 88%, SD 1.810), the answers were easy to understand (mean 88%, SD 1.732), the FAQ item was important (mean 83%, SD 1.5), and the FAQ item contained useful information (mean 90%, SD 1.536).

One of the FAQ items identified as problematic in round 1 received negative feedback concerning the terminology used to describe identifying information. Consequently, round 2 included two multiple choice questions and one open-ended question to identify the best terminology. In the round 1 FAQ, we used *identifiers* to refer to the subset of information within a record that is used in the record linkage process to distinguish one person from another, or conversely, confirm that 2 records pertain to the same individual. We also used the term *nonidentifiers* or *health-related study information* to refer to the information that researchers use to answer the research question. This terminology was critical to participants’ understanding of central concepts in the FAQ, namely, (1) researchers sometimes need identifying information such as names (eg, in record linkage), (2) identifying information is usually not needed when answering central research questions, and (3) researchers sometimes take steps to separate identifying information from the main research data before analysis (eg, coding data and using the MiNDFIRL software). There were several existing terms that were rejected as problematic either because they were terms of art with existing definitions that were not suited to a template FAQ for use in different research projects (eg, *protected health information*) or they were commonly used to refer to an entire record rather than subsets of the record (eg, *personally identifying information*).

In round 2, we asked participants for their preferred terms and included 4 options for each term, including some options suggested by participants in round 1. A plurality of round 2 participants voted for *identifying information* (16/38, 43%) and a majority of round 2 participants separately voted for *nonidentifying information* (23/37, 62%). An open-ended follow-up question allowed participants to offer additional thoughts on the suggested terms. A total of 4 participants offered support for *identifying information* or *nonidentifying information* as an alternative for their votes (Table 2).

Round 2 also contained a series of questions to improve the FAQ item pertaining to data misuse, which received the strongest negative feedback of the round 1 FAQ items. Creating a template FAQ language was particularly challenging given that the applicable laws, institutional policies, and ethical requirements are likely to vary significantly depending on the specific project [34]. Nonetheless, strong participant dissatisfaction with a vague reference to legal and institutional requirements clearly indicated that participants wanted more information. In round 2, we included 5 questions seeking feedback on what type of information participants would want for this FAQ item. Most participants agreed or strongly agreed that it was important to include links to legal requirements (30/37, 81%) or organizational rules (24/37, 65%) and thought it was important to summarize legal breach notification requirements (28/37, 76%), organizational notification requirements (25/37, 68%), or IRB required notification requirements (29/37, 78%).

Round 2 contained additional questions evaluating the revisions to the 6 other FAQ items deemed problematic in round 1 and the 3 newly added FAQ items. Participant feedback in round 2 indicated that only 3 of the FAQ items deemed problematic in round 1 were still problematic after the round 2 revisions (ie, at least three participants disagreed or strongly disagreed that the FAQ question or answer was easy to understand, important, or contained useful information). One of the open-ended comments suggested that 1 of the FAQ items remained problematic because it lacked a clear definition for *linked data*. Moreover, 1 FAQ item had conflicting feedback, specifically the need for detail versus brevity. The open-ended feedback suggested that the third FAQ item, “What difference will my data make?” was problematic for a variety of reasons, including FAQ clarity and apparent misalignment between the questions and answers included in the FAQ. In addition, several participants objected to the use of the phrase *people like you* when describing the representativeness of a sample.

Table 2. Terminology preferences of the participants for identifying and nonidentifying information (n=38).

Terminology preferences	Values, n (%)
Subset of information used to link records (ie, identifying information)[a]	
Identifiers	3 (8)
Identifying information	16 (43)
Identifiable information	4 (11)
Information that can be used to identify individuals	14 (38)
Subset of information that is used to answer the research question (ie, nonidentifying information)[b]	
Nonidentifiers	2 (5)
Nonidentifying information	23 (62)
Nonidentifiable information	5 (14)
Health-related study data	7 (19)

[a] Other terms suggested by participants: private information, personal information, patient data, confidential data, confidential information, information for identifying individuals, and personally identifiable (identifying) information.

[b] Other terms suggested by participants: data for health studies, publicly available data, social data, information that does not reveal your identity, anonymous study data, information that does not identify individuals, and information that is used as data.

All 3 of the newly introduced FAQ items were deemed problematic based on round 2 feedback. Of these new FAQ items, 1 described record linkage in general ("What is patient matching?") and the other 2 specifically described the privacy-preserving record linkage software, MiNDFIRL. The negative open-ended feedback from these 3 questions indicated that some participants struggled with the length, complexity, and sometimes the tone (ie, patronizing) of the FAQ explanations. However, negative feedback on the length and complexity often directly conflicted with positive feedback related to detail and clarity. For example, 1 participant stated, "The technical terminology cannot be avoided. However, it is most informative. I would leave this section as is." Similarly, although many participants indicated that the included visuals were helpful, the feedback indicated that the visuals were a point of confusion and frustration to others.

One new FAQ item ("How can MiNDFIRL help patient-matching while hiding identifiers?") used a text narrative to describe how a researcher would use MiNDFIRL to conduct record linkage. Participants found this static text narrative clear, but tedious to follow. In round 3, the substance of this FAQ item was incorporated into a YouTube video demonstration (ie, FAQ text as the script) of the MiNDFIRL software [35].

Delphi Round 3

A total of 33 (33/37, 89%) participants were included in round 3. We included 1 round 3 question that specifically solicited feedback on the preference of the participants for simplicity and brevity versus detail and completeness, (ie, whether we should eliminate 2 explanatory subsections of the FAQ item, "What pieces of information about me will the researchers see?"). A strong majority (25/33, 76%) of the participants preferred that the subsections be included for those who wanted the information.

On the basis of the feedback from rounds 1 and 2, we included a question in round 3 to solicit feedback on how to communicate statistics to the patient audience. The statistic described the proportion in percentage of identifying information that was revealed to a user of a MiNDFIRL software prototype as opposed to a user linking records without the software. Feedback from prior rounds suggested that patients would struggle to understand percentages. However, a majority of participants seemed to favor using percentages when asked in round 3.

In round 3, we provided alternative options for 2 problematic FAQ items. A majority of the participants (19/33, 58% and 20/33, 61%) favored the shorter alternative of each FAQ.

One FAQ item that was deemed problematic in round 2 was converted into a YouTube video demonstration for round 3 (Multimedia Appendix 8) [35]. Participants overwhelmingly found this video helpful or very helpful (31/33, 94%). Negative feedback was minor and focused on production, sound quality, and technical issues with playing the video.

In total, 3 of the 6 remaining problematic FAQ items had conflicting feedback in round 2. These items received negative feedback from a minority of our participants (10%-16%). For these questions, conflicting participant feedback in round 2 (ie, negative feedback vs positive feedback) raised doubts about substantial revisions to appease the minority of participants with negative feedback being able to maximize participant scores further. For example, we received positive feedback on the level of detail and completeness of these FAQ items and negative feedback requesting that the same FAQ items be shorter. We were unsure how to address these divergent comments with text edits. Thus, instead of adding specific questions in round 3, we revised these FAQ items where participant feedback indicated clear areas of improvement, for example, correcting visuals and adding requested definitions.

In addition, we attempted to address conflicting participant feedback on the competing values of simplicity and brevity versus detail and completeness by changing the format of the FAQ document in a few ways. First, we created an interactive website FAQ with expandable FAQ sections. This permitted users to quickly access sections that interest them and not be overwhelmed by the volume of other content. Second, we cut most definitions from the main text of the FAQ and replaced them with definition pop-up boxes that appear when a user's mouse hovers over a key term. This reduced the volume of text while still providing information if needed. Third, we bolded the text of important information within each FAQ section to aid content skimming. A strong majority of the round 3 participants (31/33, 94%) found this revised format helpful or very helpful.

We asked our expert panel one 5-point Likert question on the overall helpfulness of the FAQ website. In total, 30 of the 33 respondents rated the FAQ website as *very helpful* (n=21), *helpful* (n=6), or *somewhat helpful* (n=3).

Preliminary Thematic Analysis

The preliminary inductive thematic analysis identified 9 themes in the open-ended responses of the participants for all 3 rounds. These themes are summarized in Textbox 1. Representative quotes for these themes are available in Multimedia Appendix 9.

Textbox 1. Preliminary themes identified by inductive analysis and a brief description.

Simplicity and brevity

- Expressing a preference for short and direct explanations

Detail and completeness

- Expressing a preference for complete explanations with sufficient details for clarity and understanding

Readability

- Expressing a preference for content that is easy to read and uses layman language

Terminology and definitions

- Expressing a preference for clearly defined terms and avoiding technical jargon

Tone

- Concerning the tone of explanations (eg, conversational and not patronizing)

Examples

- Concerning the utility of examples of key concepts

Visuals

- Concerning the utility of graphics, video, and interactive aids

Data disposition and future uses

- Expressing concerns relating to what happens to research data at the end of a project (eg, destruction, reuse, and storage)

Patient rights

- Concerning the explanation of patient rights and protection

Participant Attrition

There were no significant participant attrition issues. Of the 38 participants who completed round 1, only 6 did not complete all 3 rounds. All 6 were women (aged 27-57 years) who reported having acquired a bachelor's, master's, or doctoral degree and reported fair to excellent health. Moreover, 4 of these participants were White.

Discussion

Principal Findings

Although the primary aim of this study is to develop a template language that could be used in a privacy statement for research using a specific software (MiNDFIRL), our results provide broad insights into communication with patients regarding how their data are used in research and how different software is used to enhance privacy. Communicating effectively with patients is an essential component of public health ethics, yet it is exceedingly difficult to communicate effectively [15,36]. Secondary database research uses an exceedingly broad range of research aims and methodologies. Moreover, the public is increasingly concerned about how their data are being used by the government, academia, and businesses [19,37]. Thus, proper transparency requires clearly describing the technical subject matter, communicating the purpose and process of data use, and addressing common concerns while using language that can be understood by a lay audience. To the best of our knowledge, this is the first effort to develop documentation in patient voice on an open-source software to facilitate transparent communication with the public on issues related to how a software addresses privacy concerns.

Our results indicate several considerations when communicating with patients. These considerations align with existing knowledge of health literacy (eg, simple language, visual aids, define terms, and cultural considerations) [38].

Although it may be obvious, our results support the notion that the patient community has varied preferences and opinions. This diversity manifested itself in comments about various aspects of the template disclosure language. For instance, there were frequent disagreements about the competing virtues of comprehensiveness versus conciseness, which might relate to varied levels of health literacy within our sample (ie, patients with chronic conditions) [38].

Communicating with the correct voice or tone when discussing highly technical subject matter can also be challenging. For instance, some of our participants expressed strong preferences for definitions of key terms. However, 1 respondent conveyed that an FAQ section with several definitions came across as a *textbook* answer. We used the second-person tense in our FAQ, which received positive feedback when used correctly (ie, personal and conversational). On the basis of participant feedback, we believe that a personal and conversational tone can be effective in cultivating trust between researchers and the public. However, we discovered that the second-person tense can come across as patronizing or worse. For example, when discussing representativeness, we used the phrase *people like you*. Although some participants had no issues with this phrase, several participants thought that the phrase had a negative tone and connotation and should be used carefully, if used at all.

Similarly, our results show that providing examples can prove to be beneficial and can create new challenges. Some participants found the examples to be helpful or requested examples for certain content. However, adding examples increases the content volume for the readers to absorb. If designed well, visuals can be particularly helpful to some patients. During the brief turnaround between Delphi rounds, we developed imperfect visuals that were helpful to many, but were confusing to others. Participant feedback was critical for identifying and correcting potentially confusing elements in the visuals.

The results from round 1 allow for another interesting observation. In round 1, we provided participants with a noninteractive FAQ document (PDF) and solicited feedback from participants as they progressed through each section of the document. The feedback from 8 of the 38 participants on earlier questions suggested that readers were developing questions and identifying concerns as they read the document. Although many of these questions and concerns were not germane to the section that they reviewed (and usually covered in detail subsequently), the feedback suggests that readers will either have preexisting questions that they want to be answered or will develop new questions as they read a privacy statement. This provides some support for an FAQ-style privacy statement, which identifies common questions and provides answers, as opposed to traditional privacy statements that can be challenging to navigate for personally relevant content [39].

The web-based, interactive FAQ received strong support. The interactive features (eg, expandable text and on-demand definitions) allowed us to provide information to those who want it, while reducing on-screen text for those who feel overwhelmed. This finding is consistent with other research showing preferences for web-based approaches in some populations [40].

We note that some organizations are moving away from web-based FAQs to communicate issues with the public [41]. For example, related FAQ answers might contain duplicative content, contributing to longer documents, and using questions as headings (as opposed to their answers) might slow a reader’s search for information. Our template FAQ included measures to control some of these concerns. For example, on-demand content for key terms and concepts reduces visible duplicative content, and highlighted key points aid readers in finding important information. However, future research should explore whether alternative modes of research transparency in secondary database research are preferable to an FAQ.

In addition to the FAQ-specific feedback, our participants identified a number of concerns that may be common in the broader patient community. For example, participants were interested in the use of encryption, the use of which seemed to relieve some concerns. Similar to prior work [33], patients also expressed concerns regarding the use of specific identifiers (ie, names and social security number). Several participants expressed questions and concerns relating to any future use, maintenance, or destruction of data pertaining to them. Researchers developing a privacy statement or disclosure for future research should consider including information on these topics to address these common concerns.

Limitations

This study has several limitations. First, we set out to develop a template privacy statement language to be adapted for secondary database research projects using a specific record linkage software, MiNDFIRL. As a result, our template language may have limited applicability outside of secondary database research involving record linkage. Second, our *expert* patient Delphi panel may not broadly represent the diversity of patient experiences. We note that although our Delphi panel was mostly White, female, and well educated, patients who did not complete the study had similar characteristics. Although we observed saturation in the responses indicating that the coverage of issues was good among the participant group, future work should closely consider demographic differences, particularly for those populations that have been historically marginalized in research. For example, we conducted a subsequent survey of the FAQ with a nationally representative sample (ie, meeting US Census demographic targets) of more than 500 people. Initial results indicate that, in general, more than 80% were moderately, very, or extremely satisfied, but future work should focus on investigating the differential preferences by demographics and socioeconomic status, including education level, to better detect gaps. Third, the use of the Delphi method may have resulted in repeated exposure bias over the 3 rounds. Finally, the Delphi method is useful for identifying consensus, but it cannot measure effectiveness. Future evaluations should include pre-post testing to assess the effectiveness of this template language.

Comparisons With Prior Work

There has been much work on understanding the experiences and perspectives of participants, especially in the context of using health data for research [5,9,10,17,26,40]. However, this study is the first to employ a Delphi method to engage patients with the objective of creating a privacy statement in an FAQ format to communicate issues, risks, and benefits of using their data and a software in a record linkage study.

Conclusions

Although this study aimed to develop a communication tool for use with a specific record linkage prototype software (ie, MiNDFIRL), the lessons of this research have broad applicability to efforts of researchers to communicate with data subjects.

Our results support the existence of a diverse patient population with varied preferences and information needs. Despite these diverse preferences and needs, we were able to develop a consensus language that can help communicate complex research issues, including identifiability, record linkage, and technical privacy protections. We believe that this patient-friendly language is adaptable to other research contexts.

Moreover, our findings support a compromise between individuals who want detail and individuals who want brevity. An interactive FAQ document helps connect patients with the answers that they care about and enables on-demand additional content (ie, definitions, additional explanation, and examples) without cluttering the page for all readers. Adopting patient-friendly public disclosures relating to privacy safeguards and risks, such as the template FAQ, will help to promote transparency and trust in researchers among the public and the patient community.

Acknowledgments

This research was supported in part by the Patient-Centered Outcomes Research Institute (PCORI; ME-1602-34486). PCORI did not review or approve this manuscript before publication.

Authors' Contributions

CS, HK, and AF were responsible for the conceptualization, study design and implementation, data analysis, interpretation, writing draft and revisions, and supervision. KA, TG, and GI were responsible for the study implementation, data analysis, and writing draft and revisions, and WN was responsible for the conceptualization, study design, interpretation, and revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Delphi round 1 survey instrument.
[PDF File (Adobe PDF File), 477 KB - jmir_v22i12e20783_app1.pdf]

Multimedia Appendix 2
Delphi round 1 summary provided to round 2 participants.
[PDF File (Adobe PDF File), 770 KB - jmir_v22i12e20783_app2.pdf]

Multimedia Appendix 3
Delphi round 2 survey instrument.
[PDF File (Adobe PDF File), 1289 KB - jmir_v22i12e20783_app3.pdf]

Multimedia Appendix 4
Delphi round 2 summary provided to round 3 participants.
[PDF File (Adobe PDF File), 749 KB - jmir_v22i12e20783_app4.pdf]

Multimedia Appendix 5
Delphi round 3 survey instrument.
[PDF File (Adobe PDF File), 308 KB - jmir_v22i12e20783_app5.pdf]

Multimedia Appendix 6
Delphi round 3 summary provided to Delphi participants.
[PDF File (Adobe PDF File), 271 KB - jmir_v22i12e20783_app6.pdf]

Multimedia Appendix 7
Final web-based frequently asked question template language.
[PDF File (Adobe PDF File), 2926 KB - jmir_v22i12e20783_app7.pdf]

Multimedia Appendix 8
Video demonstration of Minimum Necessary Disclosure For Interactive Record Linkage provided to Delphi participants.
[MP4 File (MP4 Video), 3796 KB - jmir_v22i12e20783_app8.mp4]

Multimedia Appendix 9
Thematic analysis with representative patient quotes.
[DOCX File , 15 KB - jmir_v22i12e20783_app9.docx]

References

1. McLeod S. The Milgram Experiment. Simply Psychology. 2017. URL: https://www.simplypsychology.org/milgram.html [accessed 2020-01-21]
2. Sterling RL. Genetic research among the Havasupai--a cautionary tale. Virtual Mentor 2011 Feb 1;13(2):113-117 [FREE Full text] [doi: 10.1001/virtualmentor.2011.13.2.hlaw1-1102] [Medline: 23121851]
3. Spitz V. Doctors from Hell: The Horrific Account of Nazi Experiments on Humans. Boulder, CO: Sentient Publications; 1996.
4. Corbie-Smith G. The continuing legacy of the Tuskegee syphilis study: considerations for clinical investigation. Am J Med Sci 1999 Jan;317(1):5-8. [doi: 10.1097/00000441-199901000-00002] [Medline: 9892266]
5. Damschroder LJ, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, privacy and trust: patients' willingness to allow researchers to access their medical records. Soc Sci Med 2007 Jan;64(1):223-235. [doi: 10.1016/j.socscimed.2006.08.045] [Medline: 17045717]
6. Federal Policy for the Protection of Human Subjects. Federal Register. 2017. URL: https://www.gpo.gov/fdsys/pkg/FR-2017-01-19/pdf/2017-01058.pdf [accessed 2017-04-14]
7. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. US Department of Health and Human Services, Office for Human Research Protections. 1979. URL: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html [accessed 2019-09-06]
8. Kum H, Krishnamurthy A, Machanavajjhala A, Ahalt SC. Social genome: putting big data to work for population informatics. Computer 2014 Jan;47(1):56-63. [doi: 10.1109/mc.2013.405]
9. Kass NE, Natowicz MR, Hull SC, Faden RR, Plantinga L, Gostin LO, et al. The use of medical records in research: what do patients want? J Law Med Ethics 2003;31(3):429-433 [FREE Full text] [doi: 10.1111/j.1748-720x.2003.tb00105.x] [Medline: 14626550]
10. O'Brien EC, Rodriguez AM, Kum H, Schanberg LE, Fitz-Randolph M, O'Brien SM, et al. Patient perspectives on the linkage of health data for research: Insights from an online patient community questionnaire. International Journal of Medical Informatics 2019 Jul;127:9-17. [doi: 10.1016/j.ijmedinf.2019.04.003]
11. Kum HC, Ragan ED, Ilangovan G, Ramezani M, Li Q, Schmit C. Enhancing Privacy through an Interactive On-demand Incremental Information Disclosure Interface: Applying Privacy-by-Design to Record Linkage. In: Fifteenth Symposium on Usable Privacy and Security. 2019 Presented at: SOUPS'19; May 1-4, 2019; Santa Clara, CA.
12. Ragan ED, Kum HC, Ilangovan G, Wang H. Balancing Privacy and Information Disclosure in Interactive Record Linkage with Visual Masking. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 9-11, 2018; New York, NY, USA. [doi: 10.1145/3173574.3173900]
13. Kum HC. Privacy Preserving Interactive Record Linkage (PPIRL) via Information Suppression: Combined Committee Meeting. 2020. Population Informatics Lab. 2020. URL: https://pinformatics.org/ppirl/pdf/all_committee202002.pdf
14. Weitzner DJ, Abelson H, Berners-Lee T, Feigenbaum J, Hendler J, Sussman GJ. Information accountability. Commun ACM 2008 Jun;51(6):82-87. [doi: 10.1145/1349026.1349043]
15. World Health Organization. WHO Guidelines on Ethical Issues in Public Health Surveillance. Geneva, Switzerland: World Health Organization; 2017.
16. Burke RD, Hammond KJ, Kulyukin V, Lytinen SL, Tomuro N, Schoenberg S. Question answering from frequently asked question filesxperiences with the FAQ finder system. AI Magazine 1997;18(2):57. [doi: 10.1609/AIMAG.V18I2.1294]
17. Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. Public opinion about the importance of privacy in biobank research. Am J Hum Genet 2009 Nov;85(5):643-654 [FREE Full text] [doi: 10.1016/j.ajhg.2009.10.002] [Medline: 19878915]
18. Smith HJ, Milberg SJ, Burke SJ. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. MIS Quarterly 1996 Jun;20(2):167. [doi: 10.2307/249477] [Medline: 249477]
19. Culnan MJ, Armstrong PK. Information privacy concerns, procedural fairness, and impersonal trust: an empirical investigation. Organ Sci 1999 Feb;10(1):104-115. [doi: 10.1287/orsc.10.1.104]

20. Foxman ER, Kilcoyne P. Information technology, marketing practice, and consumer privacy: ethical issues. J Public Policy Mark 2018 Dec 6;12(1):106-119. [doi: 10.1177/074391569501200111]
21. Cate FH. The privacy and security policy vacuum in higher education. EDUCAUSE Rev 2006;41(5):18-29.
22. Powell C. The Delphi technique: myths and realities. J Adv Nurs 2003 Feb;41(4):376-382. [doi: 10.1046/j.1365-2648.2003.02537.x] [Medline: 12581103]
23. Adler M, Ziglio E, editors. Gazing into the Oracle: The Delphi Method and its Application to Social Policy and Public Health. Philadelphia, USA: Jessica Kingsley; Jan 2, 1996.
24. Williams PL, Webb C. The Delphi technique: a methodological discussion. J Adv Nurs 1994 Jan;19(1):180-186. [doi: 10.1111/j.1365-2648.1994.tb01066.x] [Medline: 8138622]
25. Hasson F. Research guidelines for the Delphi survey technique. J Adv Nurs 2000;32(4):1008. [doi: 10.1046/j.1365-2648.2000.01567.x]
26. Hall LK, Kunz BF, Davis EV, Dawson RI, Powers RS. The cancer experience map: an approach to including the patient voice in supportive care solutions. J Med Internet Res 2015 May 28;17(5):e132 [FREE Full text] [doi: 10.2196/jmir.3652] [Medline: 26022846]
27. Albright J, de Guzman C, Acebo P, Paiva D, Faulkner M, Swanson J. Readability of patient education materials: implications for clinical practice. Appl Nurs Res 1996 Aug;9(3):139-143. [doi: 10.1016/S0897-1897(96)80254-0]
28. Badarudeen S, Sabharwal S. Assessing readability of patient education materials: current role in orthopaedics. Clin Orthop Relat Res 2010 Oct;468(10):2572-2580 [FREE Full text] [doi: 10.1007/s11999-010-1380-y] [Medline: 20496023]
29. Giannouchos T, Kum HC. Patients' And stakeholders' Perceptions of risk and benefits of the privacy preserving interactive record linkage (PPIRL) framework. New York, USA: Academy Health; 2014.
30. Hsu CC, Sandford BA. The delphi technique: making sense of consensus. Pract Assess Res Eval 2007;12(10):1-8. [doi: 10.7275/pdz9-th90]
31. Butterworth T, Bishop V. Identifying the characteristics of optimum practice: findings from a survey of practice experts in nursing, midwifery and health visiting. J Adv Nurs 1995 Jul;22(1):24-32. [doi: 10.1046/j.1365-2648.1995.22010024.x] [Medline: 7560532]
32. Population Informatics Lab Population Informatics Lab. GitHub. 2020. URL: https://github.com/pinformatics/mindfirl/tree/master/docs [accessed 2020-05-27]
33. Frequently Asked Questions. Population Informatics Lab, Texas A&M University. 2019. URL: https://pinformatics.tamhsc.edu/ppirl/faq/faq.htm [accessed 2020-05-24]
34. Schmit C, Kelly K, Bernstein J. Cross sector data sharing: necessity, challenge, and hope. J Law Med Ethics 2019 Jun;47(2_suppl):83-86. [doi: 10.1177/1073110519857325] [Medline: 31298141]
35. MINDFIRL Demo. YouTube. 2019. URL: https://www.youtube.com/watch?v=xM_Yw4h6nn4&feature=youtu.be [accessed 2020-05-26]
36. Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press; 2004.
37. Phelps J, Nowak G, Ferrell E. Privacy concerns and consumer willingness to provide personal information. J Public Policy Mark 2018 Oct 10;19(1):27-41. [doi: 10.1509/jppm.19.1.27.16941]
38. Health Literacy. Health Resources & Services Administration. 2019. URL: https://www.hrsa.gov/about/organization/bureaus/ohe/health-literacy/index.html [accessed 2020-05-25]
39. Arcand M, Nantel J, Arles - Dufour M, Vincent A. The impact of reading a web site's privacy statement on perceived control over privacy and perceived trust. Online Inform Rev 2007 Oct 2;31(5):661-681. [doi: 10.1108/14684520710832342]
40. Matar JL, Maybery DJ, McLean LA, Reupert A. Web-based health intervention for young people who have a parent with a mental illness: delphi study among potential future users. J Med Internet Res 2018 Oct 31;20(10):e10158 [FREE Full text] [doi: 10.2196/10158] [Medline: 30381281]
41. Richards S. FAQs: Why We Don't Have Them. Government Digital Service. 2013 Jul 25. URL: https://gds.blog.gov.uk/2013/07/25/faqs-why-we-dont-have-them/ [accessed 2020-09-20]

Abbreviations

FAQ: frequently asked question
IRB: institutional review board
MiNDFIRL: Minimum Necessary Disclosure For Interactive Record Linkage
NGT: nominal group technique
PCORI: Patient-Centered Outcomes Research Institute
PPRN: patient-powered research network

Edited by G Eysenbach; submitted 09.06.20; peer-reviewed by J Jose, E Whitley; comments to author 31.07.20; revised version received 24.09.20; accepted 26.10.20; published 15.12.20.

Please cite as:
Schmit C, Ajayi KV, Ferdinand AO, Giannouchos T, Ilangovan G, Nowell WB, Kum HC
Communicating With Patients About Software for Enhancing Privacy in Secondary Database Research Involving Record Linkage: Delphi Study
J Med Internet Res 2020;22(12):e20783
URL: http://www.jmir.org/2020/12/e20783/
doi:10.2196/20783
PMID:33320097

Original Paper

A Picture Is Worth a Thousand Views: A Triple Crossover Trial of Visual Abstracts to Examine Their Impact on Research Dissemination

Sandra Oska[1], BSc; Edgar Lerma[2,3], MD; Joel Topf[1], MD

[1]William Beaumont School of Medicine, Oakland University, Rochester, MI, United States
[2]College of Medicine, University of Illinois at Chicago, Chicago, IL, United States
[3]Associates in Nephrology, SC, Chicago, IL, United States

Corresponding Author:
Joel Topf, MD
William Beaumont School of Medicine
Oakland University
586 Pioneer Dr
Rochester, MI, 48309
United States
Phone: 1 248 470 8163
Email: joel.topf@gmail.com

Abstract

Background: A visual abstract is a graphic summary of a research article's question, methods, and major findings. Although they have a number of uses, visual abstracts are chiefly used to promote research articles on social media.

Objective: This study aimed to determine if the use of visual abstracts increases the visibility of nephrology research shared on Twitter.

Methods: A prospective case-control crossover study was conducted using 40 research articles published in the *American Journal of Nephrology* (AJN). Each article was shared by the AJN Twitter account in 3 formats: (1) the article citation, (2) the citation with a key figure from the article, and (3) the citation with a visual abstract. Tweets were spaced 2 weeks apart to allow washout of the previous tweet, and the order of the tweets was randomized. Dissemination was measured via retweets, views, number of link clicks, and Altmetric scores.

Results: Tweets that contained a visual abstract had more than twice as many views as citation-only tweets (1351, SD 1053 vs 639, SD 343) and nearly twice as many views as key figure tweets (1351, SD 1053 vs 732, SD 464). Visual abstract tweets had 5 times the engagements of citation-only tweets and more than 3.5 times the engagements of key figure tweets. Visual abstract tweets were also associated with greater increases in Altmetric scores as compared to citation-only tweets (2.20 vs 1.05).

Conclusions: The use of visual abstracts increased visibility of research articles on Twitter, resulting in a greater number of views, engagements, and retweets. Visual abstracts were also associated with increased Altmetric scores as compared to citation-only tweets. These findings support the broader use of visual abstracts in the scientific community. Journals should consider visual abstracts as valuable tools for research dissemination.

(J Med Internet Res 2020;22(12):e22327) doi:10.2196/22327

KEYWORDS

social media; science communication; visual abstract; Twitter; dissemination

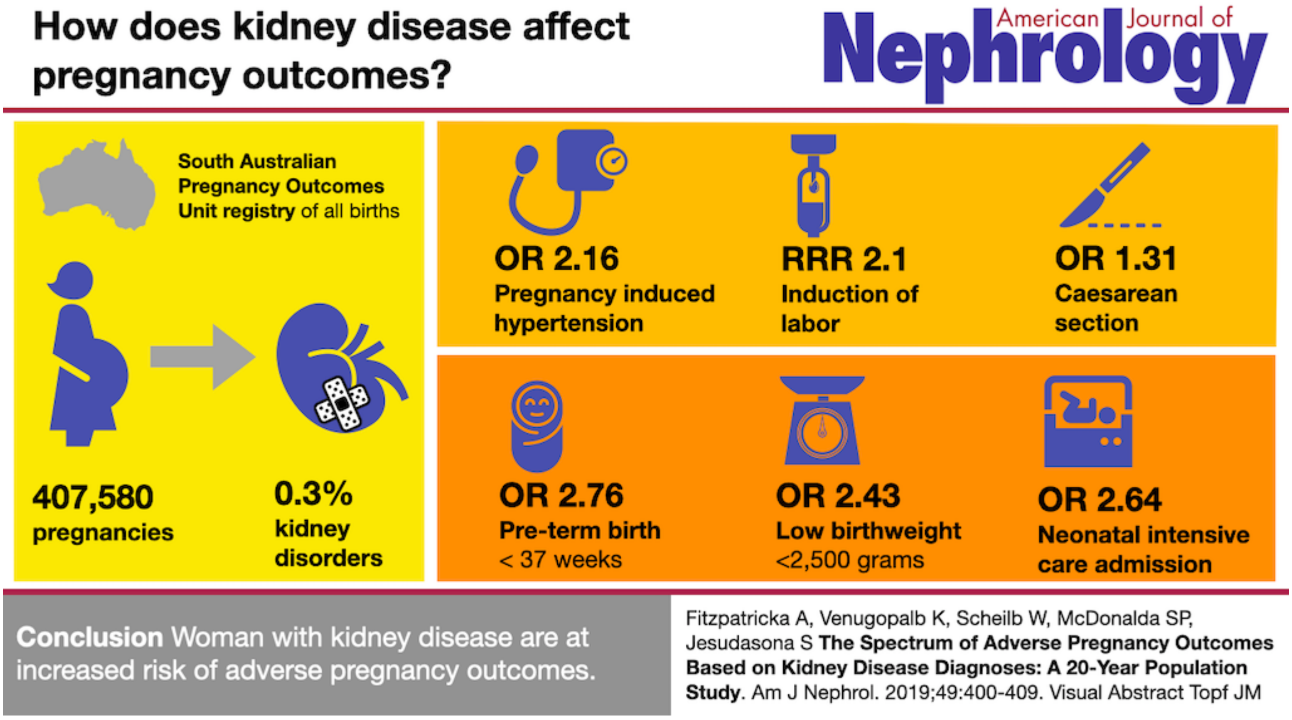
Introduction

The visual abstract is a graphic summary of a research article's question, methods, and major findings (Figure 1) [1]. Although visual abstracts were initially conceived as a way to visually represent studies for slide presentations, they were quickly adapted to promote and share studies on social media. Similar to a text abstract, a visual abstract does not replace the act of reading an article, but instead allows a viewer to quickly grasp what question a study addresses, how the investigators examined this question, and the associated results. Visual abstracts are used to accompany scientific articles on various social media platforms that rely heavily on image-based communication.

Twitter, specifically, is frequently used to share research findings and generate conversations among medical professionals and scientists [2-6]. Twitter allows physicians, journals, and institutions to disseminate research beyond the traditional readership of scientific publications to a broader audience comprised of researchers, clinicians, and the general public.

The utility and impact of the visual abstract was first evaluated by Ibrahim et al [7]. The *Annals of Surgery* shared 44 original research articles via Twitter twice each, once with only the article title and hyperlink, and once with the article title, hyperlink, and visual abstract. Ibrahim et al found that tweets with the visual abstract had a 7-fold increase in impressions (views), an 8-fold increase in retweets, and a nearly 3-fold increase in article visits [7]. Although the authors concluded that visual abstracts were responsible for the increased dissemination, it is possible that the inclusion of *any image* would have produced similar results. Our study evaluates whether visual abstracts improve viewership more than the inclusion of relevant key figures. Additionally, the visual abstract study by Ibrahim et al [7] was conducted when the use of visual abstracts was still novel. It is possible that as the use of visual abstracts has increased on social media, the associated effects have attenuated. In this study, we aim to assess the impact of visual abstract tweets as compared to key figure and citation-only tweets.

Figure 1. A sample visual abstract published alongside an American Journal of Nephrology original research publication.



Conclusion Woman with kidney disease are at increased risk of adverse pregnancy outcomes.

Fitzpatricka A, Venugopalb K, Scheilb W, McDonalda SP, Jesudasona S **The Spectrum of Adverse Pregnancy Outcomes Based on Kidney Disease Diagnoses: A 20-Year Population Study**. Am J Nephrol. 2019;49:400-409. Visual Abstract Topf JM

Methods

Overview

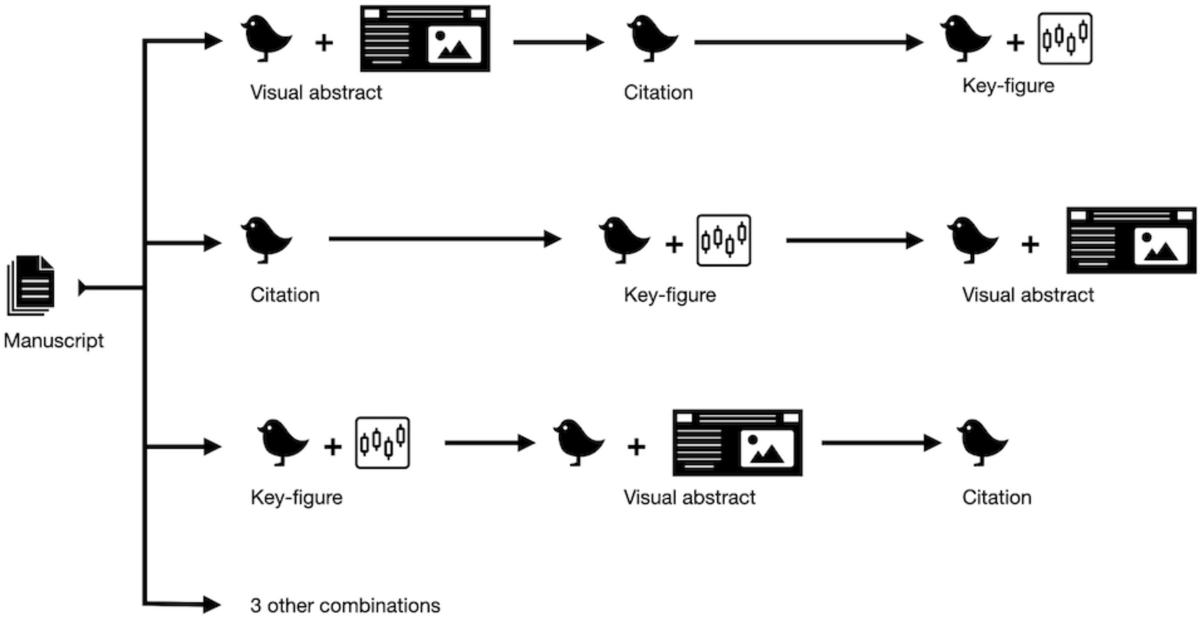
Between December 2018 and October 2019, a prospective, crossover study was conducted using original research articles published in the *American Journal of Nephrology* (AJN). Every article that had a visual abstract during the study period was included in the study. Three were excluded from analysis due to technical difficulties with the tweets or the AJN website. All research articles published in the sections "Patient-Oriented, Translational Research," "Laboratory Investigation," and "Transplantation" had visual abstracts. "Novel Research Findings" was the only section in which published research articles did not have visual abstracts. Editorials, review articles, errata, and letters did not have visual abstracts and were not included.

Each article was tweeted from the AJN Twitter account (*@AmJNephrol*) in 3 formats:

1. Citation-only
   tweets included the study title, the authors, and author institutions. If the authors or institutions had a presence on Twitter, links to those accounts were also included. All tweets had links to the original article in AJN.
2. Key figure
   tweets included the citation-only information along with a key figure from the article. The figure was chosen by one of the investigators (SO).
3. Visual abstract
   tweets included the citation-only information along with a visual abstract created by the AJN visual abstract team.

To minimize confounding from the order of tweets, each article was randomized into 1 of 6 groups. Randomization was done using the random number generator in Microsoft Excel to assign every manuscript to one of 6 groups. Each group represented the order in which the visual abstract tweet, key figure tweet, and citation-only tweet would be posted. For example, articles randomized to group 1 were tweeted in the order of visual abstract, followed by citation-only, followed by key figure (Figure 2). The rest of the groups were a permutation of this order. For each article, a 2-week “washout” period was employed between tweets. Additionally, the number of followers that the AJN account had at the time of each tweet was recorded.

Figure 2. Schematic representation of how each article was tweeted 3 times.



Assessment and Data Extraction

All analysis was performed using SAS 9.4 (SAS Institute Inc). Descriptive statistics were obtained for the outcome variables. All outcome variables were analyzed using negative binomial models. Negative binomial models were used because the data are count data, as all possible values of the variables were positive and the variables were positive integers. The engagement rate was analyzed using One-Way Analysis of Variance (ANOVA).

Altmetric scores were recorded for each study before each tweet and at the end of each washout period. Altmetric tracks the attention that research outputs receive online [8]. Altmetric scores are calculated using an automated algorithm that weighs the amount of attention a scholarly work has received in social and traditional media. The absolute difference in Altmetric scores was calculated for all tweets. Twitter analytics [9] was used to measure impressions (views), total engagements, retweets, article visits, and engagement rates for tweets. Twitter analytics defines engagements as the total number of users that interacted with a tweet. Interactions include retweets, replies, likes, following the Twitter account, clicking on a link or hashtag, playing embedded media, clicking on the author’s profile photo or name, or expanding the tweet. Link clicks represent the number of times users clicked the link in the tweet taking users to the study on the AJN website. The engagement rate is the number of engagements divided by the number of views.

Results

The study included 40 original nephrology publications. The 40 publications were randomized to Groups 1-6: Group 5 contained 5 publications (12.5%); Groups 1-4 and Group 6 had 7 publications each (17.5%). In citation-only tweets, the average number of impressions, engagements, retweets, and link clicks did not differ significantly by group (*P*=.58, *P*=.23, *P*=.77, and *P*=.32, respectively). In key figure tweets, there was no significant difference by group for average impressions, engagements, retweets, and link clicks (*P*=.07, *P*=.29, *P*=.58, and *P*=.57, respectively). For visual abstract tweets, the average number of impressions, engagements, retweets, and link clicks did not significantly differ by group (*P*=.08, *P*=.20, *P*=.64, and *P*=.37, respectively).

Visual abstract tweets had more than twice as many views as citation-only tweets (1351, SD 1053 vs 639, SD 343; *P*<.001) and nearly twice as many views as key figure tweets (1351, SD 1053 vs 732, SD 464; *P*<.001) (Table 1). Visual abstracts had 5 times the engagement of citation-only tweets (*P*<.001) and more than 3.5 times the engagement of key figure tweets (*P*<.001). Visual abstract tweets had a higher engagement rate (5.7%) than both citation-only tweets (2.4%; *P*<.001) and key figure tweets (3.2%; *P*<.001).

There was no significant difference in the number of link clicks for visual abstract tweets and key figure tweets (*P*=.35), or for visual abstract tweets and citation-only tweets (*P*=.35). The increase in Altmetric scores was significantly higher (*P*=.02) for the 2 weeks following a visual abstract tweet than for the 2 weeks following a citation-only tweet. There was no significant difference in Altmetric score for the 2 weeks following a key figure tweet compared to the 2 weeks following a visual abstract tweet (P=.85).

Table 1. Comparison of measures between tweet types.

	Citation-only tweet	Key figure tweet	VA[a] tweet	*P* value for VA vs citation only	*P* value for VA vs key figure
Number of impressions	638.58	731.88	1351.08	<.001	<.001
Number of total engagements	15.75	24.08	83.75	<.001	<.001
Number of retweets	0.95	1.30	3.68	<.001	<.001
Number of link clicks	8.13	6.68	8.95	.35	.35
Difference in Altmetric scores	1.05	2.33	2.20	.02	.85
Engagement rate	3.18	2.46	5.75	<.001	<.001

[a] Visual abstract.

Discussion

Overview

Visual abstracts increase the visibility of scientific publications on Twitter. In our crossover trial, the inclusion of a visual abstract roughly doubled the number of Twitter accounts that saw a tweet compared to a tweet without any image or compared to a tweet with a key figure from the article. However, views are passive. In order to assess how visual abstracts encourage more active forms of viewership, we tracked tweet engagement levels. Here, visual abstract tweets were even more impactful than both citation-only and key figure tweets, with visual abstracts increasing Twitter engagement by 5 fold and 3.5 fold, respectively. Although others have documented the effect of a visual abstract on research dissemination, this trial is, to our knowledge, the first to control for the effect of an image in the tweet. This is important because images in general increase views and engagement [10]. In order for visual abstracts to be a meaningful part of a dissemination strategy, they have to be more effective than a key figure from the article. The number of link clicks, however, did not differ significantly for visual abstract, key figure, and citation-only tweets. Although the online visibility of the articles improved with the addition of a visual abstract, the addition of a visual abstract did not cause users to click the link revealing the full article text. In the Ibrahim et al [7] crossover study, tweeting a visual abstract increased article visits by nearly three-fold; however, this still represented less than 1% of the total impressions. In fact, our link click rate of 0.6% with a visual abstract is similar to the 0.8% rate recorded by Ibrahim et al [7]. These very low link click rates solidify the message that although social media can boost article exposure, that exposure is shallow. Social media can convey a message quickly, but it rarely pulls viewers in for deeper consideration. Additionally, many viewers may be more interested in quickly viewing and evaluating the key elements of a paper rather than clicking through to the article link, which is often hidden behind a paywall.

A number of studies have assessed the effect of Twitter promotion on research impact and dissemination. In 2011, Eysenbach [11] showed that Twitter promotion was associated with future citations. A number of retrospective trials have reproduced these results [12,13]. A pair of subsequent randomized controlled trials from the same group were not able to show increased 30-day readership for articles randomized to twitter promotion [14,15]. However, in a trial of 696 cardiovascular articles randomized to twitter promotion or not, the tweeted group had a 1.4-fold increase in citations as well as increased Altmetric scores [16]. This was replicated by the Thoracic Surgery Social Media Network, which randomized 112 original scientific research articles from The Annals of Thoracic Surgery and The Journal of Thoracic and Cardiovascular Surgery to a tweeted group and a control group [17]. The articles were tweeted by one user and then retweeted by a team of 11 additional contributors (total audience of 52,893 followers). After 1 year of follow-up, the tweeted articles had over 4 times the number of citations as the nontweeted control group (3.1, SD 2.4 vs 0.7, SD 1.3). Of interest to our study, the authors found a high correlation with Altmetric score and citation count (R=0.72).

Several studies have demonstrated a positive association between Altmetric scores and citation rates. Azer and Azer [18] demonstrated that the Altmetric score for highly cited articles in medical professionalism literature published during or after 2007 was significantly correlated with citation counts. Similarly, Haddon Mullins et al [19] assessed the utility of Altmetric scores in general surgery literature and demonstrated that Altmetric scores had a significant positive correlation with citation number. Barbic et al [20] identified the 200 most frequently cited articles in the top 10 emergency medicine journals and concluded there was a low correlation between citation counts and Altmetric scores. The findings of such studies suggest that Altmetric scores can complement other tools and strategies that quantify research dissemination. A prospective trial by Luc et al [21] demonstrated that tweeting studies resulted in significantly more article citations over time and showed that the Altmetric score was an independent predictor of citations. In another prospective study, 24 articles were randomized to infographic or control groups and were disseminated through Twitter and Facebook [22]. Altmetric scores and abstract views were both higher in the infographic group than in the control group.

Only a few studies have looked at visual abstracts and their ability to promote articles [23]. In the previously mentioned Ibrahim et al [7] study, including a visual abstract resulted in a 7-fold increase in impressions, an 8-fold increase in retweets, and a nearly 3-fold increase in article visits. Thoma et al [24] assessed how various social strategies, including the use of infographics and podcasts, were able to promote articles from the *Canadian Journal of Emergency Medicine*. Although the study was not randomized, the authors found that when articles were promoted on social media using infographics, abstract views more than doubled and Altmetric scores nearly tripled . In a follow-up randomized controlled trial, the same group found that when articles were tweeted with infographics, abstract views and article reads doubled compared to abstract-only tweets [22]. Lindquist et al [25] demonstrated an over 6-fold increase in views in their N-of-1 experience of using a visual abstract for a geriatric medicine article.

Many journals are adopting social media strategies to promote the articles they publish [26]. However, simple strategies such as just tweeting the names of and links to articles have failed to improve readership or dissemination in multiple randomized controlled trials [15,27]. As social media becomes noisier, more sophisticated strategies are needed to rise above the din. Visual abstracts are part of this strategy and have been adopted by 95 journals. It takes humans 6 seconds to read 20 words, but the meaning of a visual symbol can be established in a quarter of a second [1]. This means that people casually scrolling through a social media feed can quickly glean the meaning of visual abstract, while a longer, text-based description may be ignored. Additionally, visual abstracts are turning up outside of social media feeds. Visual abstracts are regularly found in lectures and presentations at regional and national meetings, further promoting the article and the publishing journal [23].

Principal Results

To our knowledge, this is the first study to demonstrate that the presence of a visual abstract uniquely boosts views and improves engagement. This demonstrates that the use of a visual abstract is more valuable than a relevant image from the article. The crossover nature of our study design limited the effect of multiple graphic designers with variable talent levels in creating visual abstracts. Additionally, each article was randomized to 1 of 6 groups. Each group corresponded to the order in which the visual abstract, figure, and text would be tweeted. This randomization was employed to minimize the effect of order on impressions, total engagements, engagement rate, retweets, link clicks, and Altmetric scores. As mentioned in the results, there was no significant difference between impressions, total engagements, engagement rate, retweets, or link clicks by group. Although the AJN twitter account gained 560 followers during the time of the study, the number of followers at the time of tweeting did not differ significantly by tweet type. The triple crossover design employed in our study allows articles to be compared to themselves and removes concerns that more authors, higher budgets, or home institutions with more engagement in promoting their faculty could alter the results.

Limitations

This study does have a number of limitations. We only used Twitter analytics and Altmetric to assess dissemination of articles. As such, we cannot comment on whether the journal site was visited more frequently or whether the use of visual abstracts was associated with article downloads, an important finding in the landmark study by Ibrahim et al [7]. Because all studies were tweeted with and without a visual abstract, our study cannot determine if visual abstracts increase subsequent citations. Two of the authors (JT, EL) have large social media footprints, and their tweeting may have swayed the results. Ideally, their use of Twitter in regard to these articles should have been protocolized (either to tweet all of them or none of them). Although this was discussed during the study design, no such protocol was implemented. Additionally, we did not evaluate the tweeting patterns of the authors of the included articles. If authors were active on Twitter and had large social media footprints, this may have affected the magnitude of the results. However, by using a crossover design where each article serves as its own control, we feel that we have neutralized the effect of authors and institutions on Twitter. In addition, editorials, review articles, and letters did not have visual abstracts and were therefore not included in the study. The visual abstract design is less appropriate for such article types. Our study cannot comment on the utility of visual abstracts for these publication types.

Comparison With Prior Work

While our study has some similarities to the original Ibrahim et al [7] study on visual abstracts, we found a smaller magnitude of effect. Possible explanations could be that as visual abstracts become more common, the effect size is smaller. Another possible explanation is that not all visual abstracts are created equal, and that the visual abstracts used in the landmark study [7] were better at attracting attention than the visual abstracts created for this study. Additionally, because both of these studies are single journal studies, the nature of the journal could be an unaddressed confounder. *The Annals of Surgery* is the official journal of the American Surgical Association, while AJN is a journal not associated with a professional society.

Conclusions

This study demonstrates that the addition of a visual abstract increases the attention that an article attracts on social media. This further establishes the visual abstract as an effective tool to improve research dissemination on social media. Although a visual abstract is not a substitute for reading an article, the adoption of visual abstracts by journals can enhance the online visibility of original research. The effect size and significance of our findings support the adoption of visual abstracts by journals hoping to improve online visibility of recently published original research.

Acknowledgments

We would like to thank the AJN for allowing us to participate in tweeting from the AJN twitter account and for allowing us to access the twitter analytics reports for the account. We would like to thank Matt Sparks and Swapnil Hiremath for critical help with the paper. We would also like to thank the AJN Visual Abstract Designers: Divya Bajpai, Aldo R Jimenez Vega, Aakash Shingada, and Omar Taco Sánchez.

Conflicts of Interest

JT is a full-time nephrologist that works for St Clair Specialty Physicians. He has participated in advisory boards for Tricida, Bayer, Cara Therapeutics, and Astra Zeneca. He produces visual abstracts for Nephron, AJN, CJASN, and Kidney 360. EL is a full-time nephrologist that works for Associates in Nephrology, SC. He has participated in advisory boards for Bayer, Tricida and Astra Zeneca. He produces visual abstracts for Nephron, AJN, CJASN and Kidney360. SO has no conflicts of interest to disclose.

References

1. Ibrahim AM. Seeing is Believing: Using Visual Abstracts to Disseminate Scientific Research. Am J Gastroenterol 2018 Apr;113(4):459-461. [doi: 10.1038/ajg.2017.268] [Medline: 28925990]
2. Lamb C, Gilbert S, Ford A. Tweet success? Scientific communication correlates with increased citations in Ecology and Conservation. PeerJ 2018;6:e4564-e4N/A [FREE Full text] [doi: 10.7717/peerj.4564] [Medline: 29666750]
3. Jeong JW, Kim MJ, Oh H, Jeong S, Kim MH, Cho JR, et al. The impact of social media on citation rates in coloproctology. Colorectal Dis 2019 Oct 19;21(10):1175-1182. [doi: 10.1111/codi.14719] [Medline: 31124259]
4. O'Glasser AY, Jaffe RC, Brooks M. To Tweet or Not to Tweet, That Is the Question. Semin Nephrol 2020 May;40(3):249-263. [doi: 10.1016/j.semnephrol.2020.04.003] [Medline: 32560773]
5. Chan W, Leung A. Use of Social Network Sites for Communication Among Health Professionals: Systematic Review. J Med Internet Res 2018 Mar 28;20(3):e117 [FREE Full text] [doi: 10.2196/jmir.8382] [Medline: 29592845]
6. Rolls K, Hansen M, Jackson D, Elliott D. How Health Care Professionals Use Social Media to Create Virtual Communities: An Integrative Review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
7. Ibrahim AM, Lillemoe KD, Klingensmith ME, Dimick JB. Visual Abstracts to Disseminate Research on Social Media: A Prospective, Case-control Crossover Study. Ann Surg 2017 Dec;266(6):e46-e48. [doi: 10.1097/SLA.0000000000002277] [Medline: 28448382]
8. Kwok R. Research impact: Altmetrics make their mark. Nature 2013 Aug 22;500(7463):491-493. [doi: 10.1038/nj7463-491a] [Medline: 23977678]
9. Twitter Analytics. Twitter. URL: https://analytics.twitter.com/
10. 9 Best Social Media and Content Marketing Tips From Buffer Internet. Widrich L. 2020 Jul 10. URL: https://www.convinceandconvert.com/content-marketing/9-best-from-buffer/
11. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011 Dec 19;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
12. de Winter JCF. The relationship between tweets, citations, and article views for PLOS ONE articles. Scientometrics 2014 Sep 26;102(2):1773-1779. [doi: 10.1007/s11192-014-1445-x]
13. Smith ZL, Chiang AL, Bowman D, Wallace MB. Longitudinal relationship between social media activity and article citations in the journal Gastrointestinal Endoscopy. Gastrointest Endosc 2019 Jul;90(1):77-83. [doi: 10.1016/j.gie.2019.03.028] [Medline: 30935934]
14. Fox CS, Gurary EB, Ryan J, Bonaca M, Barry K, Loscalzo J, et al. Randomized Controlled Trial of Social Media: Effect of Increased Intensity of the Intervention. J Am Heart Assoc 2016 Apr 27;5(5):1-8 [FREE Full text] [doi: 10.1161/JAHA.115.003088] [Medline: 27121850]
15. Fox CS, Bonaca MA, Ryan JJ, Massaro JM, Barry K, Loscalzo J. A Randomized Trial of Social Media From. Circulation 2015 Jan 06;131(1):28-33. [doi: 10.1161/circulationaha.114.013509]
16. Ladeiras-Lopes R, Clarke S, Vidal-Perez R, Alexander M, Lüscher TF. Twitter promotion predicts citation rates of cardiovascular articles: a preliminary analysis from the ESC Journals Randomized Study. Eur Heart J 2020 Sep 07;41(34):3222-3225. [doi: 10.1093/eurheartj/ehaa211] [Medline: 32306033]
17. Luc JG, Archer MA, Arora RC, Bender EM, Blitz A, Cooke DT, et al. Does Tweeting Improve Citations? One-Year Results from the TSSMN Prospective Randomized Trial. Ann Thorac Surg 2020 Jun 03:1-1. [doi: 10.1016/j.athoracsur.2020.04.065] [Medline: 32504611]
18. Azer SA, Azer S. Top-cited articles in medical professionalism: a bibliometric analysis versus altmetric scores. BMJ Open 2019 Jul 31;9(7):e029433. [doi: 10.1136/bmjopen-2019-029433] [Medline: 31371297]
19. Mullins CH, Boyd CJ, Corey BL. Examining the Correlation Between Altmetric Score and Citations in the General Surgery Literature. J Surg Res 2020 Apr;248:159-164. [doi: 10.1016/j.jss.2019.11.008] [Medline: 31901796]
20. Barbic D, Tubman M, Lam H, Barbic S. An Analysis of Altmetrics in Emergency Medicine. Acad Emerg Med 2016 Mar 29;23(3):251-268 [FREE Full text] [doi: 10.1111/acem.12898] [Medline: 26743680]

21. Luc JG, Archer MA, Arora RC, Bender EM, Blitz A, Cooke DT, et al. Does Tweeting Improve Citations? One-Year Results from the TSSMN Prospective Randomized Trial. Ann Thorac Surg 2020 Jun 03:1-1. [doi: 10.1016/j.athoracsur.2020.04.065] [Medline: 32504611]
22. Huang S, Martin L, Yeh C, Chin A, Murray H, Sanderson W, et al. The effect of an infographic promotion on research dissemination and readership: A randomized controlled trial. Canadian Journal of Emergency Medicine 2018;20(6):826-833. [doi: 10.1017/cem.2018.436] [Medline: 30289098]
23. Ramos E, Concepcion B. Visual Abstracts: Redesigning the Landscape of Research Dissemination. Seminars in Nephrology 2020;40(3):291-297. [Medline: 32560778]
24. Thoma B, Murray H, Huang SYM, Milne WK, Martin LJ, Bond CM, et al. The impact of social media promotion with infographics and podcasts on research dissemination and readership. CJEM 2017 Sep 13;20(2):300-306. [doi: 10.1017/cem.2017.394] [Medline: 28899440]
25. Lindquist LA, Ramirez‐Zohfeld V. Visual Abstracts to Disseminate Geriatrics Research Through Social Media. J Am Geriatr Soc 2019 Mar 15:1128-1131. [doi: 10.1111/jgs.15853] [Medline: 30875100]
26. Lopez M, Chan T, Thoma B, Arora V, Trueger N. The Social Media Editor at Medical Journals: Responsibilities, Goals, Barriers, and Facilitators. Academic Medicine 2019;94(5):701-707. [doi: 10.1097/acm.0000000000002496] [Medline: 30334841]
27. Hawkins C, Hunter M, Kolenic G, Carlos R. Social Media and Peer-Reviewed Medical Journal Readership: A Randomized Prospective Controlled Trial. Journal of the American College of Radiology 2017;14(5):596-602. [doi: 10.1016/j.jacr.2016.12.024] [Medline: 28268163]

Abbreviations

AJN: American Journal of Nephrology

Edited by G Eysenbach; submitted 08.07.20; peer-reviewed by A Ibrahim, M Alvarez de Mon; comments to author 29.08.20; revised version received 20.09.20; accepted 29.09.20; published 04.12.20.

Please cite as:
Oska S, Lerma E, Topf J
A Picture Is Worth a Thousand Views: A Triple Crossover Trial of Visual Abstracts to Examine Their Impact on Research Dissemination
J Med Internet Res 2020;22(12):e22327
URL: https://www.jmir.org/2020/12/e22327
doi:10.2196/22327
PMID:33275112

Original Paper

Deep Neural Network for Reducing the Screening Workload in Systematic Reviews for Clinical Guidelines: Algorithm Validation Study

Tomohide Yamada[1,2*], MD, PhD; Daisuke Yoneoka[3*], PhD; Yuta Hiraike[4], MD, PhD; Kimihiro Hino[5], PhD; Hiroyoshi Toyoshiba[5], PhD; Akira Shishido[5]; Hisashi Noma[6], PhD; Nobuhiro Shojima[2], MD, PhD; Toshimasa Yamauchi[2], MD, PhD

[1]University Institute for Population Health, King’s College London, London, United Kingdom
[2]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, University of Tokyo, Tokyo, Japan
[3]Graduate School of Public Health, St Luke's International University, Tokyo, Japan
[4]Department of Cell Biology, Harvard Medical School, Boston, MA, United States
[5]FRONTEO Healthcare Inc, Tokyo, Japan
[6]Department of Data Science, The Institute of Statistical Mathematics, Tokyo, Japan
[*]these authors contributed equally

Corresponding Author:
Tomohide Yamada, MD, PhD
University Institute for Population Health
King’s College London
Addison House
Guys Campus
London, SE1 1UL
United Kingdom
Phone: 44 (0)20 7848 6625
Email: bqx07367@yahoo.co.jp

Abstract

Background: Performing systematic reviews is a time-consuming and resource-intensive process.

Objective: We investigated whether a machine learning system could perform systematic reviews more efficiently.

Methods: All systematic reviews and meta-analyses of interventional randomized controlled trials cited in recent clinical guidelines from the American Diabetes Association, American College of Cardiology, American Heart Association (2 guidelines), and American Stroke Association were assessed. After reproducing the primary screening data set according to the published search strategy of each, we extracted correct articles (those actually reviewed) and incorrect articles (those not reviewed) from the data set. These 2 sets of articles were used to train a neural network–based artificial intelligence engine (Concept Encoder, Fronteo Inc). The primary endpoint was work saved over sampling at 95% recall (WSS@95%).

Results: Among 145 candidate reviews of randomized controlled trials, 8 reviews fulfilled the inclusion criteria. For these 8 reviews, the machine learning system significantly reduced the literature screening workload by at least 6-fold versus that of manual screening based on WSS@95%. When machine learning was initiated using 2 correct articles that were randomly selected by a researcher, a 10-fold reduction in workload was achieved versus that of manual screening based on the WSS@95% value, with high sensitivity for eligible studies. The area under the receiver operating characteristic curve increased dramatically every time the algorithm learned a correct article.

Conclusions: Concept Encoder achieved a 10-fold reduction of the screening workload for systematic review after learning from 2 randomly selected studies on the target topic. However, few meta-analyses of randomized controlled trials were included. Concept Encoder could facilitate the acquisition of evidence for clinical guidelines.

(J Med Internet Res 2020;22(12):e22422) doi:10.2196/22422

KEYWORDS

machine learning; evidence-based medicine; systematic review; meta-analysis; clinical guideline; deep learning; neural network

Introduction

Evidence-based medicine aims to provide treatment that matches a patient's needs by integrating the best and latest scientific evidence and clinical skills [1]. Performing systematic reviews and meta-analyses is vital to obtain data that can inform evidence-based clinical decisions as well as the development of clinical and public health guidelines [2].

When performing a systematic review, it is critical to minimize potential bias by identifying all relevant published articles through exhaustive and systematic screening of the literature, which can be an extremely time-consuming and resource-intensive process.

The Cochrane collaboration mandates reinvestigation and updating of published systematic reviews and meta-analyses every 2 years to maintain the novelty and quality of evidence [3], but this is an onerous task. As a single systematic review or meta-analysis usually requires 1 to 2 years to complete, only one-third of all Cochrane reviews are updated on time [4], and many reviews are obsolete or missing [5,6]. Therefore, the development of methods for the automation of the systematic review process has been suggested [7].

To reduce the time and cost of screening literature when performing systematic reviews, researchers have explored the use of active learning text classification systems to achieve semiautomated exclusion of irrelevant studies while retaining a high proportion of eligible studies for subsequent manual review [8,9]. However, little progress has been made for the following reasons. First, previous studies did not investigate well-characterized and high-quality data sets, so the type of systematic review used as the data source was unclear, and the method of applying machine learning to the clinical studies was obscure. Second, previous reports did not specify how active machine learning was used. Third, only an approximate 30%-50% reduction of the workload was achieved [8]. Fourth, a method that extracts 100% of the correct articles from the literature has not been developed because most studies use a targeted extraction of 95% as the primary outcome; despite the importance of not missing any eligible studies when performing systematic reviews (ie, the objective is to identify all relevant articles) [10-14].

To overcome some of these issues, we studied systematic reviews of randomized controlled trials cited in several recent international clinical guidelines to investigate whether an active machine learning system (Concept Encoder, Fronteo Inc) could reduce the workload and accelerate the review process while improving its precision.

Methods

Search Strategy and Selection of Reviews

This study was performed according to a specified protocol and was registered with the University Hospital Medical Information Network clinical trials registry (UMIN000032663). Our institutional review board waived the need for approval. Three reviewers (TYamada, HT, and NS) independently checked the reference lists of 5 recent clinical guidelines released by the American Diabetes Association [15], American College of Cardiology [16], American Heart Association (2 guidelines) [17,18], and American Stroke Association [19]. The reviewers identified all systematic reviews and meta-analyses cited in these guidelines with no language restrictions.

Next, the reviewers selected eligible systematic reviews and meta-analyses of interventional randomized controlled trials for medications that fulfilled the following inclusion criteria: First, a reproducible search strategy was required; therefore, articles with no description of the search strategy, or without a clear, reproducible description of the search strategy were excluded. In addition, meta-analyses using individual data, meta-analyses of observational studies, reports missing relevant information, and reviews of fewer than 5 studies were excluded. Finally, reviews were excluded if the primary screening data set did not include all of the correct articles (ie, those cited) when it was reproduced according to the published search strategy. Disagreements among the reviewers were resolved by consensus.

We reproduced primary screening data sets, including abstracts, according to reported search strategies, that is, a search strategy for PubMed was devised based on the search strategy for Ovid MEDLINE described in each review (Multimedia Appendix 1).

Active Machine Learning System

An artificial intelligence engine (Concept Encoder) [20] was used to convert sentences into vectors, extract and learn each vector component as a feature value, identify similar vectors as indicators of the similarity of sentence content, and perform a rapid search for similar sentences. Vectorization facilitates text analysis by providing numerical data that allow various calculations to be performed (eg, to assess clustering of results). In addition, vectorization allows searches to be based on the sums and differences of sentences, facilitating comparison of content between 2 sentences and resulting in a sentence retrieval engine that can be adapted to research targets.

First, each sentence is decomposed into morphemes (the smallest meaningful units of a language) by morphological analysis, applying rules to label each morpheme level element with a word. Next, the word labels were embedded in the *k*-dimension vector space [21-24] using the word2vec technique. Sentences can also be embedded in the *k*-dimension vector space using an expansion to the word-embedding method called doc2vec that yields paragraph vectors [21-24]. Several parameters are used in these embedding techniques, such as the number of embedded words, the vectors' dimensions, and negative sampling (ie, the number noise samples, nonobserved data, generated in both word2vec and doc2vec algorithms). These algorithms enable the transformation into vectors of words and documents from articles in a systematic review. Assuming that there are a total of *m* abstracts and *n* words in all the articles (both reviewed and not reviewed) in a single systematic review or meta-analysis (ie, 1 of the 8 systematic reviews or meta-analyses included) embedded in a *k*-dimension vector space, then the abstracts and words can be expressed as

Embedded vectors are well known to possess interesting features such as word analogy and outperform the bag of word approaches in several linguistic tasks. For example, if 2 articles have similar contents, then the 2 row vectors in D associated with those articles are a short cosine distance from one another. Similarly, the 2 row vectors in W associated with 2 words having a similar meaning are also a short cosine distance from one another. Hence, if there are differences between the articles that were reviewed and not reviewed, then the reviewed articles should be closer to each other than those that were not reviewed. These features persist after the 2 matrices are multiplied due to linearity of multiplication. For example, if $w_i \cong w_j$ for 2 row vectors in matrix W, then the inner product with d (a row vector in matrix D) is $d \cdot w_i \cong d \cdot w_j$. Expanding this to word analogy, if $w_i - w_i \cong w_{i'} - w_{j'}$ where $i, j, i', j' \in [1, 2, 3, ..., n]$ holds for 4 row vectors in matrix W, then $d \cdot w_i - d \cdot w_j \cong d \cdot w_{i'} - d \cdot w_{j'}$ is true for any row vector d in matrix D.

Hence, the product of these 2 matrices is a DW matrix, which is a sentence-word matrix that also possesses these interesting features of the original matrices.

In this study, sentence similarity was evaluated by using a DW matrix.

Neural networks have previously been used to calculate D and W matrices, but calculation of these matrices becomes computationally intense when a large number of articles are investigated [21-24]. Hence, a neural network is generally restricted to embedding the 1000 most common words in m articles. In our analysis, the 1000 most common words were identified for each of the 8 studies.

A skip-gram model with negative sampling was chosen to calculate W. The embedding vector dimension was set at k=300, which is usually considered sufficient to capture word and document features, and the number of negative samplings was set at n_s=5. A previous paper [25] reported that values of negative sampling in the range of n_s=5-20 were useful for small training data sets, whereas for large data sets, n_s can be as small as 2-5; the size of the data sets used in this study ranged from m=138 to m=6935.

For D, the distributed bag of words version of paragraph vector [20-24] was used as it is usually consistent across many tasks [24]. The same negative sampling and embedding dimension (n_s=5 and k=300) were used. Both D and W were obtained at the same time in this study. However, it is possible to obtain W first and then calculate D by using the pretrained W. We used the gensim (version 3.8.3)[26] package for Python (version 3.6) with n_s=5, k=300, and 1000 words.

A dimension reduction technique, such as singular value decomposition, can be used to approximate the DW matrix with a lower dimension matrix to reduce computational requirements; however, this was not done in this analysis (the number of columns in the DW matrix kept as 1000).

Reproduction of the Reviews

The similarity of any 2 articles is defined as the cosine distance of the 2 vectors associated with these articles. After a correct (reviewed) or incorrect (not reviewed) article is identified, the associated row vector is defined as a correct or incorrect and used as the feature vector representing a correct or incorrect article. The cosine distances for all other articles ($m - 1$ articles) are calculated and arranged in descending order. For the next article from the top of the list, if the article is a correct one, the mean of the vectors for the correct articles is used to train Concept Encoder in the next step of active learning. If the article is an incorrect one, the vector is subtracted to train Concept Encoder in the next step of active learning, that is, it is used as the feature vector. Cosine distances between the updated vector and all other articles are calculated and ordered again, and this process is repeated until all of the correct articles have been identified. Here, the mean vector is simply used as the feature vector for the correct articles. We could build classification models using these vectors as features to arrange the remaining articles in a descending manner by active learning; however, similarity of articles seemed to be embedded in the vectors, and using the vectors directly as the features was effective. Therefore, we kept the process simple, and no further machine learning was conducted in our active learning process.

Workload Reduction

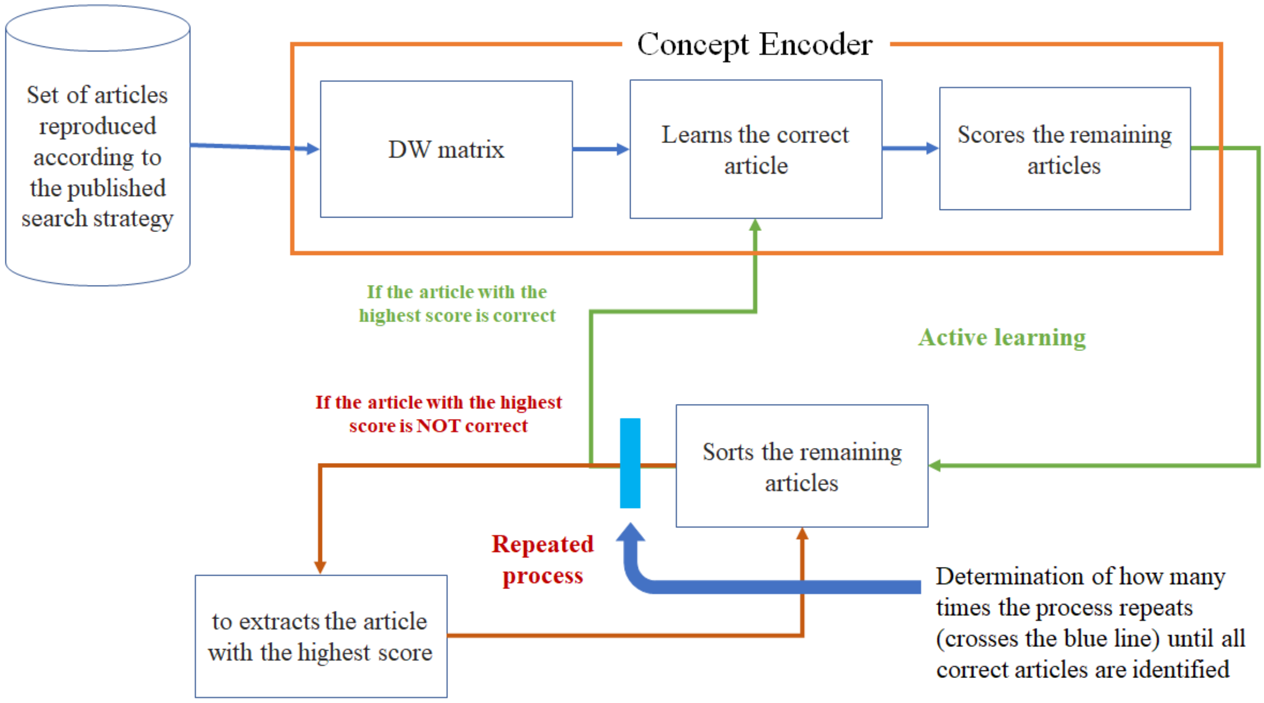
Using 2 randomly selected correct articles (selected by Concept Encoder from among the correct articles), the following steps were performed to calculate how much workload reduction could be achieved using Concept Encoder.

1. Concept Encoder read these 2 articles and calculated the mean value of the sentence-word vectors corresponding to the 2 articles. Next, this mean value was used to assign scores to the other articles by determining the cosine distance between the mean value and the vectors corresponding to each of the remaining articles (Figure 1).
2. A researcher reviewed the article with the higher score. If this was a correct article, Concept Encoder learned it as a correct article based on the mean value of all chosen sentence-word vectors. If it was an incorrect article, the sentence-word vector is subtracted from the mean vector of the corrected articles.
3. Concept Encoder learned the correct and incorrect article, and thus identified and rescored the remaining articles, which had not been checked by the researcher.
4. The researcher again reviewed the article with the highest score. If this was a correct article, Concept Encoder learned it as a correct article. If it was incorrect, Concept Encoder learned it as an incorrect article.
5. After learning all of the correct and incorrect articles identified up to this point, Concept Encoder scored the remaining articles again. The mean of sentence-word vectors for all corrected articles minus the mean of sentence-word vectors for all incorrect articles was used to score the remaining articles.
6. Steps 2 to 5 were repeated until all of the correct articles had been identified. Following this, the final reading ratio was calculated as the number of articles read by Concept

Encoder relative to the total number of articles. For example, if the total data set comprised 1000 articles, and Concept Encoder found all of the correct articles after reading 200 articles, the final reading ratio would be 20%, and the work involved in screening the literature would have been reduced by 80% (avoiding the need to read 800 out of 1000 articles). Work saved over sampling (WSS) @*R*% is an index to measure how much work is saved compared to manual screening to achieve identification of *R*% of correct papers.

7. Next, the first correct article (step 2) was changed, and the same process was repeated until all of the correct articles were identified.
8. The maximum reduction of the literature screening workload achieved by teaching Concept Encoder 2 correct articles (ie, 2 articles that were actually reviewed) was determined.

Figure 1. Flow diagram of information processing and user interaction with Concept Encoder.



Endpoints

The primary endpoint of this study was the reduction in the literature screening workload when Concept Encoder was used to identify all of the correct articles, relative to the workload for finding all of the correct articles by manual review with random sampling. WSS@95% recall was used for comparability as this endpoint is often used in previous studies (Multimedia Appendix 1).

Statistical Analysis

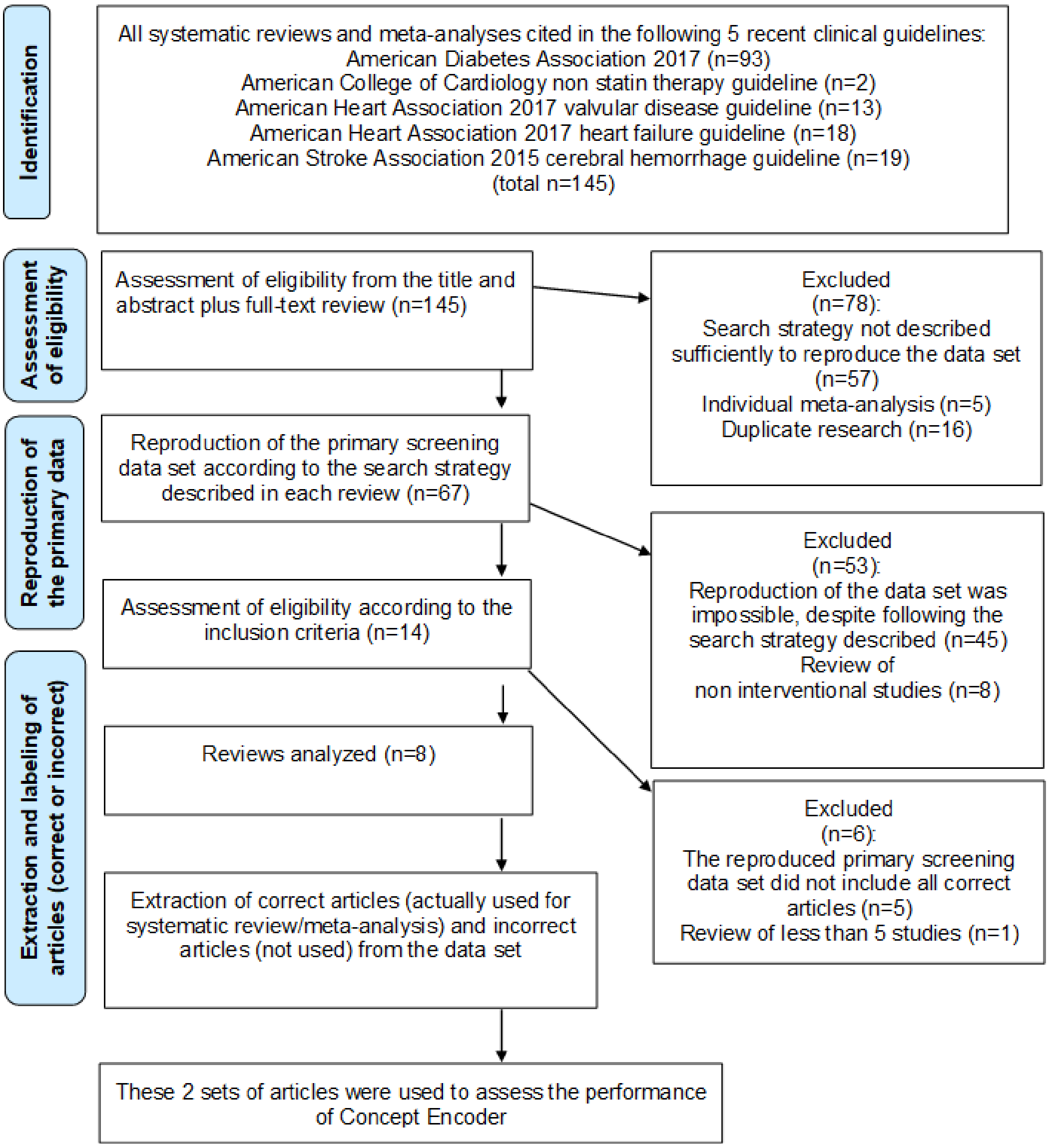
WSS and receiver operating characteristics were used to evaluate the performance of the algorithm. Area under the receiver operating characteristic curve (AUROC) shows how much the active learning improves classification ability between correct and incorrect articles at each step of learning .

To evaluate the impact of the 2 initial papers selected on system performance, all possible pairs of papers were generated and used to run the algorithm. Then the mean and standard deviation of WSS@95% were measured. The confidence interval of the AUROC was determined at each step of the active learning process for all 8 studies using scores calculated from the cosine distances for articles that were used or not used in the systematic reviews.

Results

A flowchart of our strategy for performing the literature search and study selection is shown in Figure 2. The systematic reviews and meta-analyses used in this study were cited in 5 recent clinical guidelines (93 from American Diabetes Association 2017 guidelines [15], 2 from American College of Cardiology guidelines for nonstatin therapy [16], 13 from American Heart Association 2017 guidelines for valvular disease [17], 18 from American Heart Association 2017 guidelines for heart failure [18], 19 from American Stroke Association 2015 guidelines [19]). Among the 145 candidate reviews, 137 were excluded, with the main reasons being that the search strategy was not described in sufficient detail to reproduce the data set (57 reviews), or the data set could not be reproduced despite following the described search strategy (45 reviews). A final 8 reviews published between 2012 and 2016 were selected [27-34]. These reviews comprised 2 Cochrane Database Systematic Reviews and 1 each published in JAMA Neurology, the British Medical Journal, PLOS Medicine, the Journal of the American Medical Association, the Lancet, and the Archives of Internal Medicine. The characteristics of these reviews are summarized in Multimedia Appendix 1.

Figure 2. Literature search and study selection strategy.



After reproducing the primary screening data set (including abstracts) according to the search strategy described in each review, 81 sets of correct articles and 22,664 sets of incorrect articles were obtained (Multimedia Appendix 1). The search strategies employed for the reproduction of the data sets are detailed in Multimedia Appendix 1.

One of the 8 studies contained only 140 articles. The number of words appearing more than twice in the data set was approximately 1200, including the stopwords. We also wished to examine the difference in performance between studies. Figure 3 displays the average cumulative recall curves for the 8 reviews. The performance of Concept Encoder was evaluated for every possible pair of articles chosen at the start of active learning. Concept Encoder was found to significantly reduce the workload by at least 0.867 compared with manual screening (the lowest mean WSS@95%). The average reduction of the workload compared with manual screening was >90% or 10-fold (WSS@95%: mean 0.904), and Concept Encoder showed a high ability to discriminate between correct and incorrect studies (Table 1). The choice of the initial 2 articles only had a small influence on the performance of the learning algorithm.

Prioritization (ie, the score based on cosine distance) of the algorithm by machine learning increased the AUROC to between 0.99 to 1.00, while the standard deviation of the AUROC decreased with each prioritization step (Figure 4).

Figure 3. Average cumulative recall curves for all data sets: (a) Chatterjee et al [27], (b) Balsells et al [28], (c) Muduliar et al [29], (d) Yanovski and Yanovski [30], (e) Eng et al [31], (f) McBrien et al [32], (g) Andrade Castetllanos et al [33], and (h) Arguedas et al [34]. WSS: work saved over sampling.

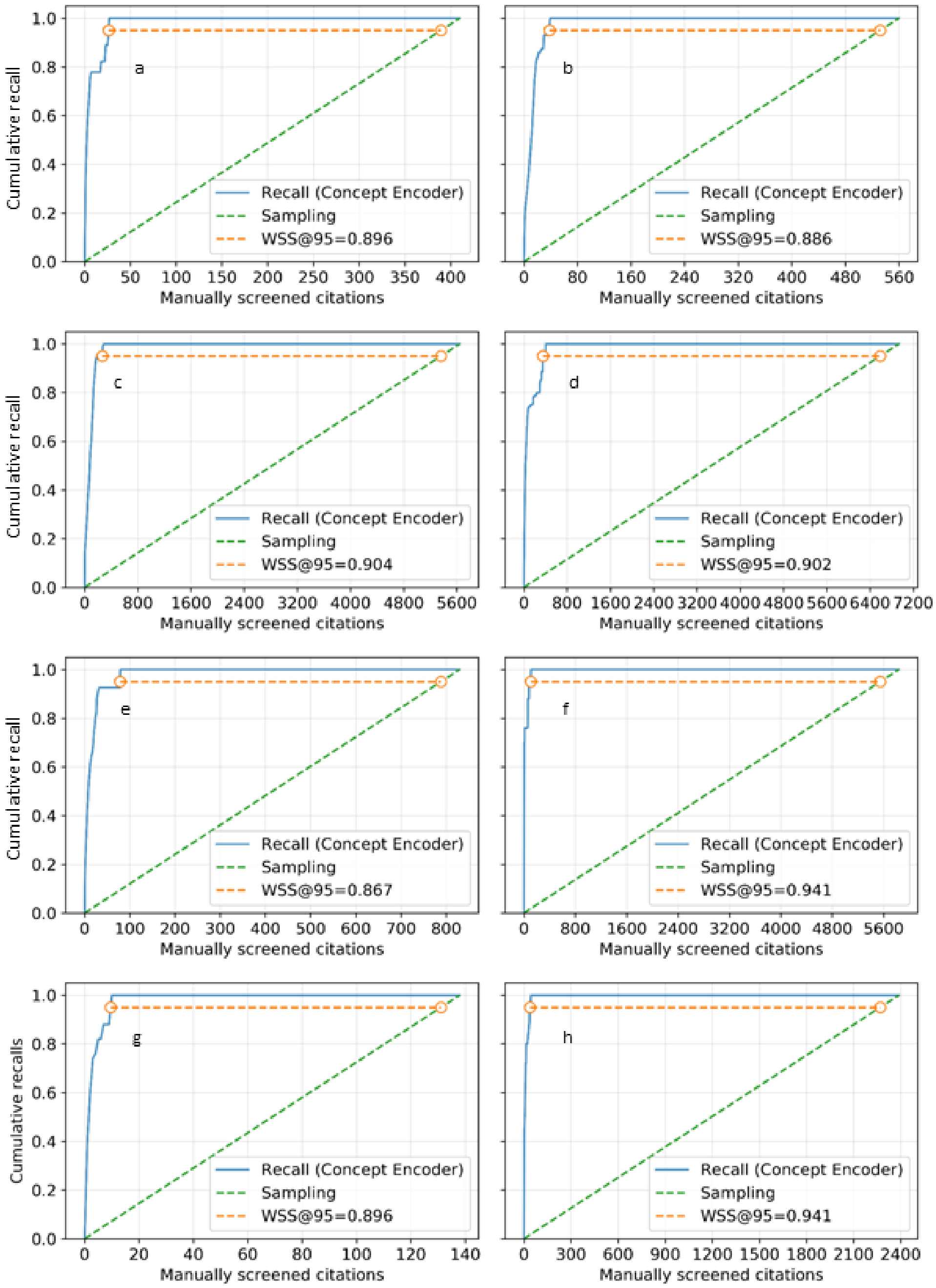


Table 1. Review data sets and corresponding results.

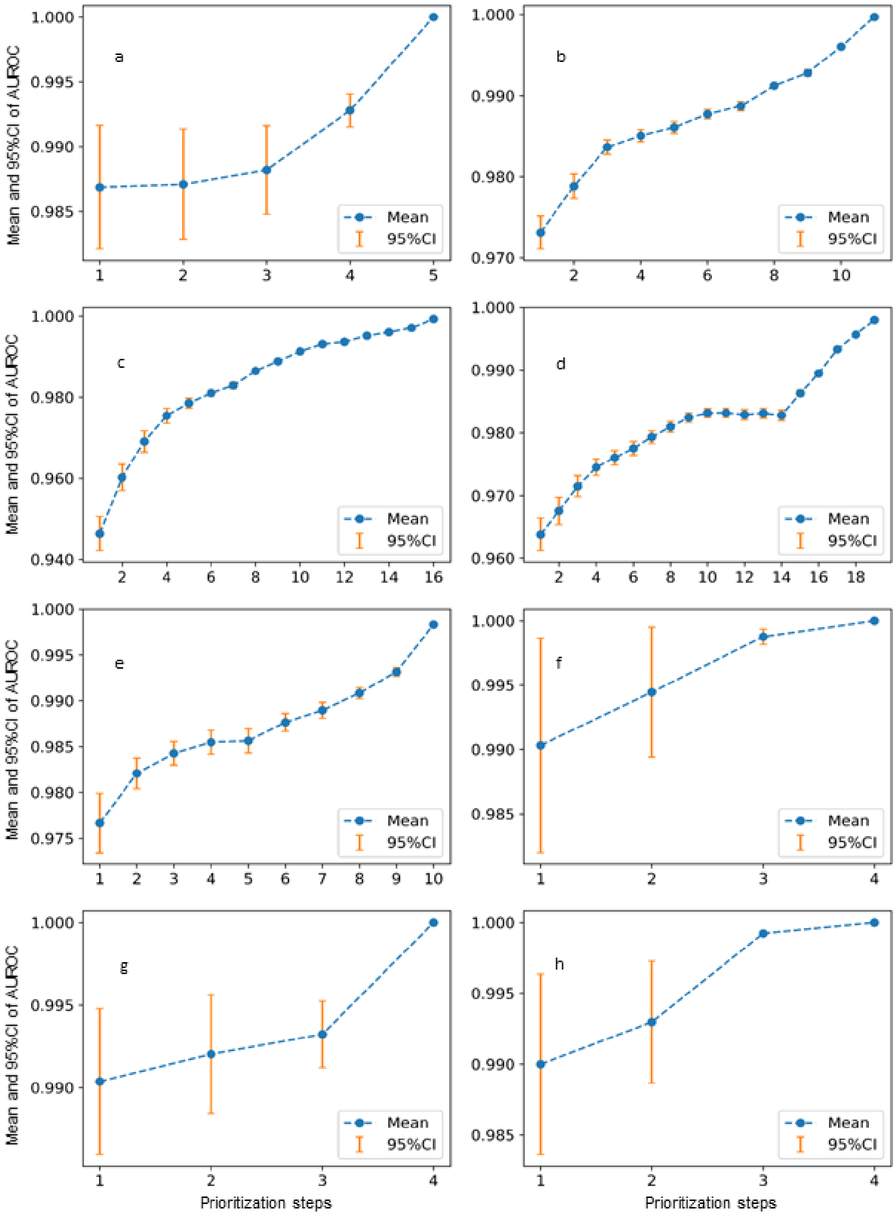
Review			Concept Encoder					
Reference	Correct articles, n	Articles screened, n	Trials, n	AUROC[a]	WSS@100[b]		WSS@95[c]	
					Mean (SD)	Range	Mean (SD)	Range
[27]	6	410	15	1	0.946 (0.014)	0.937-0.985	0.896 (0.014)	0.887-0.935
[28]	12	560	66	1	0.936 (0.009)	0.930-0.959	0.886 (0.009)	0.880-0.909
[29]	17	5644	136	0.999	0.954 (0.006)	0.946-0.971	0.904 (0.006)	0.896-0.921
[30]	20	6935	190	0.998	0.944 (0.005)	0.941-0.957	0.902 (0.006)	0.897-0.932
[31]	11	830	55	0.998	0.917 (0.023)	0.906-0.975	0.867 (0.023)	0.856-0.925
[32]	5	5839	10	1	0.991 (0.006)	0.981-0.999	0.941 (0.006)	0.931-0.949
[33]	5	138	10	1	0.946 (0.015)	0.935-0.978	0.896 (0.015)	0.885-0.928
[34]	5	2389	10	1	0.991 (0.006)	0.982-0.996	0.941 (0.006)	0.932-0.946
Mean	10	2843	62	0.999	0.953 (0.011)	0.945-0.977	0.904 (0.011)	0.895-0.931

[a] AUROC: area under the receiver operating characteristic curve.

[b] WSS@100: work saved over sampling at 100%.

[c] WSS@95: work saved over sampling at 95%.

Figure 4. Performance for an increasing number of prioritization steps: (a) Chatterjee et al [27], (b) Balsells et al [28], (c) Muduliar et al [29], (d) Yanovski and Yanovski [30], (e) Eng et al [31], (f) McBrien et al [32], (g) Andrade Castetllanos et al [33], and (h) Arguedas et al [34]. AUROC: area under the receiver operating characteristic curve.



Discussion

Principal Results

These findings demonstrated that an active machine learning system could dramatically reduce the workload for performing systematic reviews of randomized controlled trials in several medical fields. Our data suggest that an active machine learning system could improve the precision of the systematic review process as well as reduce the time required, thus assisting with the development of clinical guidelines. In this study, the deep neural network–based active machine learning system achieved a 10-fold reduction in the literature screening workload for systematic reviews after a researcher initiated the learning process by randomly selecting 2 studies.

Strengths and Limitations

We demonstrated that a 90% reduction in the workload for searching literature compared with manual assessment could be achieved and, whereas previous research mainly focused on small databases, we showed that this reduction in workload could be applied to large data sets by using systematic reviews of clinical studies. In addition, we specifically described the methods employed by our active machine learning system for systematic reviews of literature, which most previous reports do not explain.

One of the limitations of our study was the absence of a criterion for when active learning can be stopped. The study focused on how much workload could be reduced by the embedding-based technique using WSS@95%; however, active learning could increase the AUROC value as active learning steps proceeded; and therefore, at some point, this method could separate correct articles from incorrect articles in the learning process. The other limitation of the study was that 2 correct articles were required at the beginning of active learning. In practice, it may be challenging to start the review process with 2 correct articles already identified. This limitation might be overcome by using 2 consecutive systematic reviews on the same topic; the papers used in the first review could be used as the learning data to identify new articles for the second systematic review.

Comparison With Prior Work

Several studies using text mining or computational technique to reduce workload in systematic reviews have been reported. Marshall et al [35] used an ensemble model consisting of support vector classification and convolutional neural networks to classify randomized controlled trial papers and showed that the model predicted randomized controlled trial papers (AUROC 0.987, 95% CI 0.984-0.989) and also discussed the automating risk of bias assessment using large corpus labeled by distant supervision and presented a step toward automating or semiautomating the data extraction needed for the synthesis of clinical trials [36]. These authors also evaluated the performance of the RobotReviewer in another paper [37] and showed that machine learning could help reviewers to detect sentences or documents containing risk of bias but are not be able to replace manual review by humans yet. However, these works showed a great potency of workload reduction in systematic reviews with machine learning techniques. Wallace et al [38] developed a tool for systematic review called Abstrackr. Based on its technical report [38], 2 case studies were tested, and a 40% workload reduction with 100% recall was achieved. Rathbone et al [39] evaluated the performance of Abstrackr for 4 systematic reviews and summarized that reduction of workload varied from 10% to 80%, but that precision was also decreased. Recently, Gates et al [40] evaluated the Abstrackr performance retrospectively against human review for 4 studies and concluded that it could reduce workload by 9.5% to 88.4%, varying by the screening task. A review of systematic reviews [41] noted that current use of text mining in systematic reviews could reduce workload from 30% to 70%, at 95% recall. As for other techniques to reduce workload in systematic reviews, using 17 studies, RobotAnalyst [42] was reported as an active learning approach using latent Dirichlet allocation to reduce workload, for which WSS@95% varied between 6.89% to 70.74%. Workload reduction varies by study or task; therefore, direct comparison with our study is difficult. However, our method, using an embedding-based technique, showed good performance with the 8 systematic review data sets of randomized controlled trials.

Regarding other embedding methods, embedding vectors from BioBERT-Base version 1.1 (4.5 billion PubMed abstracts, trained for 1 million steps) [43] were applied to the same 8 studies. WSS@95% was calculated for each study using the same algorithm. The mean WSS@95% for the 8 studies was 0.747 (SD 0.119), which was about 15% lower than the 0.904 (SD 0.02) from this study (Table 1). Fine-tuning for each study was not performed because some of the studies include only a small number of articles. Hence, the performance of BioBERT could be improved by fine-tuning. However, the method in the present paper is still competitive enough considering the performance and simplicity of the model.

We assessed systematic reviews and meta-analyses of randomized controlled trials because these can estimate the true efficacy and risks of treatment. In contrast, estimates derived from systematic reviews and meta-analyses of epidemiological studies are more limited due to the observational design of the underlying studies. Therefore, further investigation will be needed to assess the effectiveness of our system for meta-analyses of epidemiological studies. Furthermore, in the future, we plan to evaluate Cochrane review papers, which have a standardized review process.

Conclusion

The deep neural network–based active machine learning system investigated in this study achieved at least a 10-fold reduction of the literature screening workload for systematic reviews after a researcher initiated the learning process by randomly selecting 2 studies that fulfilled the inclusion criteria for the target review. Our findings suggest that machine learning could facilitate the acquisition of evidence for developing new clinical guidelines.

Acknowledgments

TYamada was funded by The Japan Diabetes Society, Japan Society for the Promotion of Science (16 K20965), Japan Foundation for Applied Enzymology, Japan Health Promotion Foundation, Pfizer Health Research Foundation, Suzuki Manpei Diabetes Foundation, The Daiwa Anglo-Japanese Foundation, The Tanita Healthy Weight Community Trust, Daiwa Securities Health Foundation, European Foundation for the Study of Diabetes and Japan Diabetes Society Reciprocal Travel Research Fellowship Programme, The Kanae Foundation for the Promotion of Medical Science, The Takano Science Foundation, Senri Life Science Foundation, Foundation for Total Health Promotion, The Telecommunications Advancement Foundation, Okinaka Memorial Institute for Medical Research, The Uehara Memorial Foundation, Kao Healthcare Science Foundation, Fuji Foundation for Protein Research, Honjo International Scholarship Foundation, The Salt Science Research Foundation, Japan Health and Research Institute, and The Health Care Science Institute. The funding sources had no role in the design, conduct, and reporting of this study.

Authors' Contributions

TYamada, DY, KH, HT, AS, and NS conceived and designed the review. TYamada, KH, HT, AS, and NS identified reports and extracted data. TYamada, DY, YH, KH, HT, AS, and NS interpreted the data. TYamada, DY, KH, HT, and NS drafted the manuscript and all other authors (YH, AS, HN, and TYamauchi) reviewed the manuscript. All authors approved submission of this manuscript for publication. TYamada is the guarantor of this work and, as such, had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

HT, AS, and KH are employees of Fronteo Inc. The other authors declare no competing interests.

Multimedia Appendix 1
Online supplementary material.
[DOCX File , 37 KB - jmir_v22i12e22422_app1.docx]

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
2. Mulrow CD. Rationale for systematic reviews. BMJ 1994 Sep 03;309(6954):597-599 [FREE Full text] [doi: 10.1136/bmj.309.6954.597] [Medline: 8086953]
3. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions. Cochrane Training. URL: http://handbook.cochrane.org/ [accessed 2018-09-01]
4. Jaidee W, Moher D, Laopaiboon M. Time to update and quantitative changes in the results of Cochrane pregnancy and childbirth reviews. PLoS One 2010 Jul 13;5(7):e11553 [FREE Full text] [doi: 10.1371/journal.pone.0011553] [Medline: 20644625]
5. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010 Sep 21;7(9):e1000326 [FREE Full text] [doi: 10.1371/journal.pmed.1000326] [Medline: 20877712]
6. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? a survival analysis. Ann Intern Med 2007 Aug 21;147(4):224-233 [FREE Full text] [doi: 10.7326/0003-4819-147-4-200708210-00179] [Medline: 17638714]
7. Tsafnat G, Dunn A, Glasziou P, Coiera E. The automation of systematic reviews. BMJ 2013 Jan 10;346:f139. [doi: 10.1136/bmj.f139] [Medline: 23305843]
8. Wallace BC, Trikalinos TA, Lau J, Brodley C, Schmid CH. Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics 2010 Jan 26;11:55 [FREE Full text] [doi: 10.1186/1471-2105-11-55] [Medline: 20102628]
9. Wallace BC, Small K, Brodley CE, Lau J, Schmid CH, Bertram L, et al. Toward modernizing the systematic review pipeline in genetics: efficient updating via data mining. Genet Med 2012 Jul;14(7):663-669 [FREE Full text] [doi: 10.1038/gim.2012.7] [Medline: 22481134]
10. Cohen AM. Performance of support-vector-machine-based classification on 15 systematic review topics evaluated with the WSS@95 measure. J Am Med Inform Assoc 2011;18(1):104; author reply 104-104; author reply 105 [FREE Full text] [doi: 10.1136/jamia.2010.008177] [Medline: 21169622]
11. Przybyła P, Brockmeier AJ, Kontonatsios G, Le Pogam M, McNaught J, von Elm E, et al. Prioritising references for systematic reviews with RobotAnalyst: a user study. Res Synth Methods 2018 Sep;9(3):470-488 [FREE Full text] [doi: 10.1002/jrsm.1311] [Medline: 29956486]
12. Hashimoto K, Kontonatsios G, Miwa M, Ananiadou S. Topic detection using paragraph vectors to support active learning in systematic reviews. J Biomed Inform 2016 Aug;62:59-65 [FREE Full text] [doi: 10.1016/j.jbi.2016.06.001] [Medline: 27293211]
13. Kontonatsios G, Brockmeier AJ, Przybyła P, McNaught J, Mu T, Goulermas JY, et al. A semi-supervised approach using label propagation to support citation screening. J Biomed Inform 2017 Aug;72:67-76 [FREE Full text] [doi: 10.1016/j.jbi.2017.06.018] [Medline: 28648605]
14. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev 2015 Jan 14;4:5 [FREE Full text] [doi: 10.1186/2046-4053-4-5] [Medline: 25588314]
15. Standard of Medical Care in Diabetes-2017. American Diabetes Association. 2017. URL: https://care.diabetesjournals.org/content/diacare/suppl/2016/12/15/40.Supplement_1.DC1/DC_40_S1_final.pdf [accessed 2020-12-22]
16. Lloyd-Jones DM, Morris PB, Ballantyne CM, Birtcher KK, Daly DD, DePalma SM, et al. 2017 focused update of the 2016 ACC Expert Consensus Decision Pathway on the role of non-statin therapies for LDL-cholesterol lowering in the management of atherosclerotic cardiovascular disease risk: a report of the American College of Cardiology Task Force on Expert Consensus Decision Pathways. J Am Coll Cardiol 2017 Oct 03;70(14):1785-1822 [FREE Full text] [doi: 10.1016/j.jacc.2017.07.745] [Medline: 28886926]
17. Nishimura RA, Otto CM, Bonow RO, Carabello BA, Erwin JP, Fleisher LA, et al. 2017 AHA/ACC focused update of the 2014 AHA/ACC guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2017 Jun 20;135(25):e1159-e1195. [doi: 10.1161/CIR.0000000000000503] [Medline: 28298458]

18. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Colvin MM, et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation 2017 Aug 08;136(6):e137-e161. [doi: 10.1161/CIR.0000000000000509] [Medline: 28455343]
19. Hemphill JC, Greenberg SM, Anderson CS, Becker K, Bendok BR, Cushman M, American Heart Association Stroke Council, Council on CardiovascularStroke Nursing, Council on Clinical Cardiology. Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2015 Jul;46(7):2032-2060. [doi: 10.1161/STR.0000000000000069] [Medline: 26022637]
20. Concept Encoder. FRONTEO. 2018 Jan. URL: https://www.fronteo.com/en/products/conceptencoder/ [accessed 2020-12-22]
21. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv. 2013 Sep 7. URL: https://arxiv.org/format/1301.3781 [accessed 2020-12-22]
22. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. 2014 Presented at: Conference on Empirical Methods in Natural Language Processing; October 24; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
23. Dai A, Olah C, Le Q. Document embedding with paragraph vectors. arXiv. 2015 Jul 29. URL: https://arxiv.org/pdf/1507.07998 [accessed 2020-12-22]
24. Le Q, Mikolov T. Distributed representations of sentences and documents. 2014 Jun 22 Presented at: 31st International Conference on Machine Learning; 2014; Bejing, China p. 1188-1196.
25. Mikolov T, Sutskever I, Chen K. Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems 2013;3111:3119 [FREE Full text]
26. Řehůřek R, Sojka P. Software framework for topic Modelling with Large Corpora. 2010 May 22 Presented at: LREC 2010 workshop New Challenges for NLP Frameworks; 2010; Valletta, Malta p. 45-50 URL: http://is.muni.cz/publication/884893/en
27. Chatterjee S, Sardar P, Biondi-Zoccai G, Kumbhani DJ. New oral anticoagulants and the risk of intracranial hemorrhage: traditional and Bayesian meta-analysis and mixed treatment comparison of randomized trials of new oral anticoagulants in atrial fibrillation. JAMA Neurol 2013 Dec;70(12):1486-1490. [doi: 10.1001/jamaneurol.2013.4021] [Medline: 24166666]
28. Balsells M, García-Patterson A, Solà I, Roqué M, Gich I, Corcoy R. Glibenclamide, metformin, and insulin for the treatment of gestational diabetes: a systematic review and meta-analysis. BMJ 2015;350:h102 [FREE Full text] [Medline: 25609400]
29. Mudaliar U, Zabetian A, Goodman M, Echouffo-Tcheugui JB, Albright AL, Gregg EW, et al. Cardiometabolic risk factor changes observed in diabetes prevention programs in US settings: a systematic review and meta-analysis. PLoS Med 2016 Jul;13(7):e1002095 [FREE Full text] [doi: 10.1371/journal.pmed.1002095] [Medline: 27459705]
30. Yanovski SZ, Yanovski JA. Long-term drug treatment for obesity: a systematic and clinical review. JAMA 2014 Jan 01;311(1):74-86 [FREE Full text] [doi: 10.1001/jama.2013.281361] [Medline: 24231879]
31. Eng C, Kramer CK, Zinman B, Retnakaran R. Glucagon-like peptide-1 receptor agonist and basal insulin combination treatment for the management of type 2 diabetes: a systematic review and meta-analysis. Lancet 2014 Dec 20;384(9961):2228-2234. [doi: 10.1016/S0140-6736(14)61335-0] [Medline: 25220191]
32. McBrien K, Rabi DM, Campbell N, Barnieh L, Clement F, Hemmelgarn BR, et al. Intensive and standard blood pressure targets in patients with type 2 diabetes mellitus: systematic review and meta-analysis. Arch Intern Med 2012 Sep 24;172(17):1296-1303. [doi: 10.1001/archinternmed.2012.3147] [Medline: 22868819]
33. Andrade-Castellanos CA, Colunga-Lozano LE, Delgado-Figueroa N, Gonzalez-Padilla DA. Subcutaneous rapid-acting insulin analogues for diabetic ketoacidosis. Cochrane Database Syst Rev 2016 Jan 21(1):CD011281. [doi: 10.1002/14651858.CD011281.pub2] [Medline: 26798030]
34. Arguedas JA, Leiva V, Wright JM. Blood pressure targets for hypertension in people with diabetes mellitus. Cochrane Database Syst Rev 2013 Oct 30(10):CD008277. [doi: 10.1002/14651858.CD008277.pub2] [Medline: 24170669]
35. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. Res Synth Methods 2018 Dec;9(4):602-614 [FREE Full text] [doi: 10.1002/jrsm.1287] [Medline: 29314757]
36. Marshall IJ, Kuiper J, Wallace BC. Automating risk of bias assessment for clinical trials. IEEE J Biomed Health Inform 2015 Jul;19(4):1406-1412. [doi: 10.1109/JBHI.2015.2431314] [Medline: 25966488]
37. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc 2016 Jan;23(1):193-201 [FREE Full text] [doi: 10.1093/jamia/ocv044] [Medline: 26104742]
38. Wallace B, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an Interactive Machine Learning System in an Evidence-Based Practice Center. Byron Wallace. URL: http://www.byronwallace.com/static/articles/wallace_ihi_2011_preprint.pdf [accessed 2020-12-15]
39. Rathbone J, Hoffmann T, Glasziou P. Faster title and abstract screening? evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. Syst Rev 2015 Jun 15;4:80 [FREE Full text] [doi: 10.1186/s13643-015-0067-6] [Medline: 26073974]
40. Gates A, Johnson C, Hartling L. Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool. Syst Rev 2018 Mar 12;7(1):45 [FREE Full text] [doi: 10.1186/s13643-018-0707-8] [Medline: 29530097]

41. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev 2015 Jan 14;4:5 [FREE Full text] [doi: 10.1186/2046-4053-4-5] [Medline: 25588314]
42. Przybyła P, Brockmeier AJ, Kontonatsios G, Le Pogam M, McNaught J, von Elm E, et al. Prioritising references for systematic reviews with RobotAnalyst: a user study. Res Synth Methods 2018 Sep;9(3):470-488 [FREE Full text] [doi: 10.1002/jrsm.1311] [Medline: 29956486]
43. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb 15;36(4):1234-1240. [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]

Abbreviations

AUROC: area under the receiver operating characteristic curve
WSS: work saved over sampling

Edited by G Eysenbach; submitted 14.07.20; peer-reviewed by Y Kondo, S Shams; comments to author 05.08.20; revised version received 10.11.20; accepted 30.11.20; published 30.12.20.

Please cite as:
Yamada T, Yoneoka D, Hiraike Y, Hino K, Toyoshiba H, Shishido A, Noma H, Shojima N, Yamauchi T
Deep Neural Network for Reducing the Screening Workload in Systematic Reviews for Clinical Guidelines: Algorithm Validation Study
J Med Internet Res 2020;22(12):e22422
URL: https://www.jmir.org/2020/12/e22422
doi:10.2196/22422
PMID:33262102

Original Paper

Engagement With a Behavior Change App for Alcohol Reduction: Data Visualization for Longitudinal Observational Study

Lauren Bell[1], MSc; Claire Garnett[2], PhD; Tianchen Qian[3], PhD; Olga Perski[2], PhD; Elizabeth Williamson[1,4], PhD; Henry WW Potts[4,5], PhD

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom
[2]Research Department of Behavioural Science and Health, University College London, London, United Kingdom
[3]Department of Statistics, University of California Irvine, Irvine, CA, United States
[4]Health Data Research UK, London, United Kingdom
[5]Institute of Health Informatics, University College London, London, United Kingdom

Corresponding Author:
Lauren Bell, MSc
Department of Medical Statistics
London School of Hygiene and Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 02076368636
Email: lauren.bell@lshtm.ac.uk

Abstract

Background: Behavior change apps can develop iteratively, where the app evolves into a complex, dynamic, or personalized intervention through cycles of research, development, and implementation. Understanding how existing users engage with an app (eg, frequency, amount, depth, and duration of use) can help guide further incremental improvements. We aim to explore how simple visualizations can provide a good understanding of temporal patterns of engagement, as usage data are often longitudinal and rich.

Objective: This study aims to visualize behavioral engagement with *Drink Less*, a behavior change app to help reduce hazardous and harmful alcohol consumption in the general adult population of the United Kingdom.

Methods: We explored behavioral engagement among 19,233 existing users of *Drink Less*. Users were included in the sample if they were from the United Kingdom; were 18 years or older; were interested in reducing their alcohol consumption; had a baseline Alcohol Use Disorders Identification Test score of 8 or above, indicative of excessive drinking; and had downloaded the app between May 17, 2017, and January 22, 2019 (615 days). Measures of when sessions begin, length of sessions, time to disengagement, and patterns of use were visualized with heat maps, timeline plots, k-modes clustering analyses, and Kaplan-Meier plots.

Results: The daily 11 AM notification is strongly associated with a change in engagement in the following hour; reduction in behavioral engagement over time, with 50.00% (9617/19,233) of users disengaging (defined as no use for 7 or more consecutive days) 22 days after download; identification of 3 distinct trajectories of use, namely engagers (4651/19,233, 24.18% of users), slow disengagers (3679/19,233, 19.13% of users), and fast disengagers (10,903/19,233, 56.68% of users); and limited depth of engagement with 85.076% (7,095,348/8,340,005) of screen views occurring within the *Self-monitoring and Feedback* module. In addition, a peak of both frequency and amount of time spent per session was observed in the evenings.

Conclusions: Visualizations play an important role in understanding engagement with behavior change apps. Here, we discuss how simple visualizations helped identify important patterns of engagement with *Drink Less*. Our visualizations of behavioral engagement suggest that the daily notification substantially impacts engagement. Furthermore, the visualizations suggest that a fixed notification policy can be effective for maintaining engagement for some users but ineffective for others. We conclude that optimizing the notification policy to target both effectiveness and engagement is a worthwhile investment. Our future goal is to both understand the causal effect of the notification on engagement and further optimize the notification policy within *Drink Less* by tailoring to contextual circumstances of individuals over time. Such tailoring will be informed from the findings of our micro-randomized trial (MRT), and these visualizations were useful in both gaining a better understanding of engagement and designing the MRT.

(J Med Internet Res 2020;22(12):e23369) doi:10.2196/23369

KEYWORDS

mobile health; behavior change; apps; digital health; data visualizations; engagement; micro-randomized trial; push notifications; just-in-time adaptive interventions

Introduction

Background

Maintaining alcohol consumption within recommended guidance is widely known to reduce one's risk of illness or injuries. Such guidance includes the recommendations of the Chief Medical Officer of the United Kingdom to limit alcohol consumption to 14 units a week and to have frequent alcohol-free days [1]. However, anyone in the general adult population who wants to reduce their hazardous or harmful alcohol consumption may face certain challenges to follow such guidance [2]. Challenges include the ease of access to alcohol and alcohol being an addictive substance. This can lead to individuals developing chronic or cyclical patterns of excessive drinking, with the personal behaviors of drinking influenced by internal or external factors [3-5]. Internal factors refer to feeling states or events in an individual's recent drinking history, such as previous drinking episodes, moods, motives, or cravings that may modify future patterns of drinking [6]. External factors are influential events that occur independently of an individual's drinking history; for example, how the risk of hazardous drinking of the general population increases during holiday periods or weekends [7,8].

Behavior change apps, sensors, and wearables offer a way of reducing hazardous alcohol consumption through real-time data capture and interventions [9-12]. Benefits of behavior change apps, that can be synchronized with sensors and wearables, include capturing an individual's dynamic history of alcohol consumption and state of mind while providing *around the clock* access to support, particularly in moments when an individual's vulnerability to hazardous drinking may increase [13].

However, a key challenge for the majority of behavior change apps is that levels of engagement remain low [14-16]. Engagement, often a mediator of effectiveness [14], is considered a multifaceted construct composed of behavioral and experiential aspects [17]. Usage data from a behavior change app provides an understanding of behavioral engagement (hereafter referred to as engagement) with the app [18]. Multiple indicators of engagement are thought to convey important information about how users interact with a given intervention, including the frequency (eg, number of log-ins), depth (eg, proportion of available modules accessed), amount (eg, time spent per log-in), and duration (eg, total number of days) of use [19].

Drink Less is a behavior change app that aims to help its users reduce hazardous and harmful alcohol consumption. The app was developed following the multiphase optimization strategy framework (comprising a preparation phase, an optimization phase, and an evaluation phase) [20-23] and the UK Medical Research Council's guidance on developing complex interventions [24-26]. The app includes 6 different theory and evidence-informed modules: normative feedback, goal setting, cognitive bias training, self-monitoring and feedback, action planning, and identity change. These modules are described in detail by Garnett et al [27]. The app sends a local daily push notification at 11 AM that asks users to "Please complete your drink diaries," to encourage self-monitoring of drinking behavior. The default 11 AM timepoint was set so as not to disturb late risers and to allow participants time to complete their morning routine; however, the notification timing could be changed by the user.

Owing to the agile nature of app development, optimization of engagement can be done through cycles of research and implementation [28]. Identifying important patterns of engagement for such optimization purposes presents various analytical challenges that visualizations can address. Visualizations have previously been helpful for analyzing a wide variety of rich data streams within public health research [29-33]. Simple visualizations, especially when complemented with clear textual descriptions, are generally recommended for identifying and comparing trends [32]. In previous digital health research, visualizations have delivered *at a glance* insight from mass volume and time-varying data, including more sophisticated displays of spatiotemporal, contextual, and event-centric outcomes [34-38]. Importantly, visualizations can provide insights into optimization that include (1) patterns of use that may boost or hinder behavior change, (2) a better understanding of temporal engagement with various components of the intervention, and (3) pathways toward personalization of the intervention.

Objectives

The aim of this paper is to explore the usefulness of simple visualizations in uncovering important temporal patterns of engagement and facilitating decision making for further intervention development. This study presents 2 key contributions to improving engagement with *Drink Less*. The first contribution, provided in the Results section, is to showcase a number of visualizations that helped us understand temporal patterns of engagement with *Drink Less*. The second contribution, provided in the Discussion section, explains how insights obtained from these visualizations informed the next stages of intervention optimization.

Methods

Data Transformation

Each visualization involved transformation of the data. Original usage data involved merging, by an anonymous user ID, a data set of baseline characteristics (age, sex, employment type, and Alcohol Use Disorders Identification Test [AUDIT] score) to a data set of time stamps of start time of use, screen views, and length (in microseconds) of use. Along with use, the actions of entering an alcohol-free day or recording units of alcohol consumed were measured.

Data

Data set 1 included 19,233 users who downloaded *Drink Less* between May 17, 2017, and January 22, 2019 (615 days). The inclusion criteria for users included having a baseline AUDIT score of 8 or above, which is indicative of excessive drinking [39]; being from the United Kingdom; being aged 18 years or above; being interested in reducing their alcohol consumption; using app versions 1.0.11 to 1.0.16; and having consented to the Privacy Notice (Multimedia Appendix 1). Screen views data are recorded automatically and downloaded via Panda scripts from *Nodechef* (a web-based platform for hosting mobile apps) using a secure https protocol. Sessions were derived from screen views using the Pandas script.

Users who downloaded the app on August 21, 2018 (n=5830), were excluded as an article on BBC News was published on this date, which endorsed the app (Garnett et al, unpublished data, 2020); thus, these users were likely to have different characteristics and engagement behavior.

Data set 2 included time stamps of 829,001 sessions and 8,169,005 screen views of the 19,233 users in data set 1. This includes 122,332 entries of alcohol-free days and 123,704 entries of alcohol drinks consumed. All use was recorded from May 17, 2017, to April 16, 2019 (699 days). As such, users had a minimum of 84 days of use measured.

To explore various engagement aspects, we developed sets of data from data sets 1 and 2 with varying engagement measures.

Set A

All use was measured from May 17, 2017, to January 22, 2019 (615 days), including date of download and time stamps of all use. This period was chosen as it reflects a time in which the content of the app was relatively stable.

Set B

Set B included all users whose use was measured in Set A, with data only over the first 30 days from download, with the measure "Did use occur on this day?" (binary, yes or no) for each user.

Measures

Log-in Sessions and Frequency of Log-Ins

A session was defined as a continuous series of screen views, with a new session defined as a new screen view after 30 min of inactivity [40]. Clearing or *swiping away* the daily notification did not register as use and was not considered as either a session or a module view. All time stamps were appropriately adjusted from Coordinated Universal Time to British Summer Time. The amount of use per log-in session was operationalized as time spent (in seconds) per session. Daily use was captured by the measure "Did use occur on this day?" (binary, yes or no) for each user for 30 days (Set B).

Drinking Diary Entry

In the self-monitoring and feedback module, users enter an alcohol-free day and the date of its occurrence, the number of alcoholic units consumed, and the date of consumption. The time stamps in which records were made was measured.

Disengagement

We defined disengagement as the first day of 7 or more consecutive days of no use after download [41]. The days between download and disengagement were derived for each user. Users who did not disengage after downloading the app were censored.

Data Visualization Methods and Analytical Techniques

We used heat maps, timeline plots, k-modes clustering, generalized estimation equations, and Kaplan-Meier plots to explore and visualize patterns of engagement with *Drink Less*. Analyses were carried out in R [42] and Stata [43]. We used the following R library packages to create the visualizations: ggplot2 for heat maps and timeline plots [44], rayshader to create the 3D animations [45], viridis for color palettes sensitive to readers with color blindness [46], Klar to perform the k-modes clustering [47], survminer for the Kaplan-Meier survival curves and number at risk table [48], gganimate to create animated plots of use over time [49], and patchwork to place graphs side by side [50]. The data visualization methods, data set and engagement measures are shown in Textbox 1.

K-modes clustering is an extension of the k-means algorithm for partitioning categorical data, which uses a general dissimilarity measure [51,52]. Within each cluster, we visualized the probability of opening the app during the day over time with 95% CI. The appropriate number of clusters was explored through the *elbow* method and *silhouette* method [53]. The *elbow* method explains the variance of the data in relation to the number of clusters and shows by how much the addition of another cluster would reduce the dissimilarity measure. The *silhouette* method shows how well each user fits into their respective cluster through 2 distance measures: separation (ie, the average distance to the closest other cluster) and compactness (ie, the average within-cluster distance) [54,55]. Kaplan-Meier plots show the estimated cumulative proportion of users engaged and the time scale is days after download [56,57].

Textbox 1. Data visualization methods, data, and engagement measures.

Set A:

- Heat maps: Total count of sessions and total amount of time spent on *Drink Less*, by hour and day of the week
- Timeline plots: Frequency and median amount of time per session
- Kaplan-Meier plots: Time to disengagement (defined as days after download followed by 7 or more consecutive days of nonuse)

Set B:

- K-modes clustering: Was the app used or not each day, over 30 days after download

To explore the association between the delivery of the notification and subsequent near-term engagement of opening the app (ie, engagement in the hour after the notification is delivered), we compared opening the app (yes or no) between the *exposed* time period (11 AM to noon) and an *unexposed* time period (10 AM to 11 AM). We estimated the association between exposure to the notification and opening of the app, which was quantified using a risk ratio. We fitted a marginal model for the outcome of opening the app by using a generalized estimating equation [58] with robust standard errors and an independent working correlation matrix. We fitted an unadjusted model and a model adjusted for the baseline covariates of the continuous variables age, days after download and baseline AUDIT score, which were all included as linear terms, and the categorical variables employment type and gender. Further models explored effect moderation by adding an interaction between exposure to the notification and (1) days after download and (2) cluster (as identified by the k-modes analysis). In the final model, we additionally allowed the association between cluster and exposure to the notification to vary linearly by day after download. Estimated risk ratios with 95% CIs and Wald test *P* values are presented. For models with interaction terms, we present risk ratios for exposure to the notification estimated at days 1, 7, and 30 after download, estimated separately for each cluster.

Results

Overview

The user characteristics are reported in Table 1. Approximately half (49.5%) of the sample were male. The mean age of users was 44 (SD 11.2) years, and the majority worked in nonmanual employment (71.7%). Just under half (46.6%) had a baseline AUDIT score indicating hazardous alcohol consumption (8 to 15, inclusive).

Table 1. User characteristics (N=19,233).

User characteristics	Participants
Sex, n (%)	
Male	9540 (49.60)
Age (years), mean (SD)	44 (11.2)
Employment type, n (%)	
Nonmanual employment	13,792 (71.71)
AUDIT[a] risk zone, n (%)	
Hazardous (8-15)	8958 (46.58)
Harmful (16-19)	3949 (20.53)
At risk of alcohol dependence (20-40)	6326 (32.89)

[a] AUDIT: Alcohol Use Disorders Identification Test.

Summative tables of use (screen views and time on app) by module are provided in Multimedia Appendix 2. It was observed that 85% of screen views occurred in the module *Self-Monitoring and Feedback*. The number of users who reported at least one alcohol-free day or at least one alcohol drink record was 61.86% (11,898/19,233) and 49.11% (9445/19,233), respectively. Over the first 30 days of use after download (derived for Set B data), the median number of sessions per user was 9, with an IQR of 2 to 28 sessions, and the median time spent per user was 24 min, with an IQR of 9 to 55 min.

Visualizations

Patterns of Frequency of Use, Length of Use, Entries of Alcohol-Free Days, and Alcohol Units Consumed

In Figure 1, both heat maps show days of the week along the x-axis and hour of the day along the y-axis. Plot A in Figure 1 shows the frequency of opening the app by hour of the day and day of the week. This shows that there is a strong association between delivery of the notification and opening of the app in the following hour, and this is consistent throughout the week. Plot B in Figure 1 shows the amount of use by hour of the day and day of the week. This shows that the notification is also associated with the distribution of the total time spent on the app. In plot B, hotspots are observed across the evenings and on Saturday, Sunday, and Monday mornings, which are not evident in plot A. A heat map of when *Drink Less* was downloaded (Multimedia Appendix 3) shows hotspots of downloads on Sunday and Monday evenings. Rotating 3D heat map films of Figure 1, which show the variations more clearly, are provided in Multimedia Appendices 4 and 5.

In Figure 2, plot C shows the median time spent on the app along the y-axis and plot D shows the total number of sessions starting in the hour along the y-axis. Timeline plots show the hour of the day on the x-axis. Plot D shows that the frequency of sessions sharply peaks in the hour after the notification is sent at 11 AM. A second natural peak of frequency occurred in the evenings and a third smaller peak in the mornings. Plot C shows that the median length of time drastically dropped from 11 AM onward, with a slow and steady recovery as the day progressed. An animation of plot D over time is provided in Multimedia Appendix 6, showing that the shape of the distribution over 30 days remains consistent.

Figure 1. Heat maps of total frequency of use (sessions) and total time on app (hours).

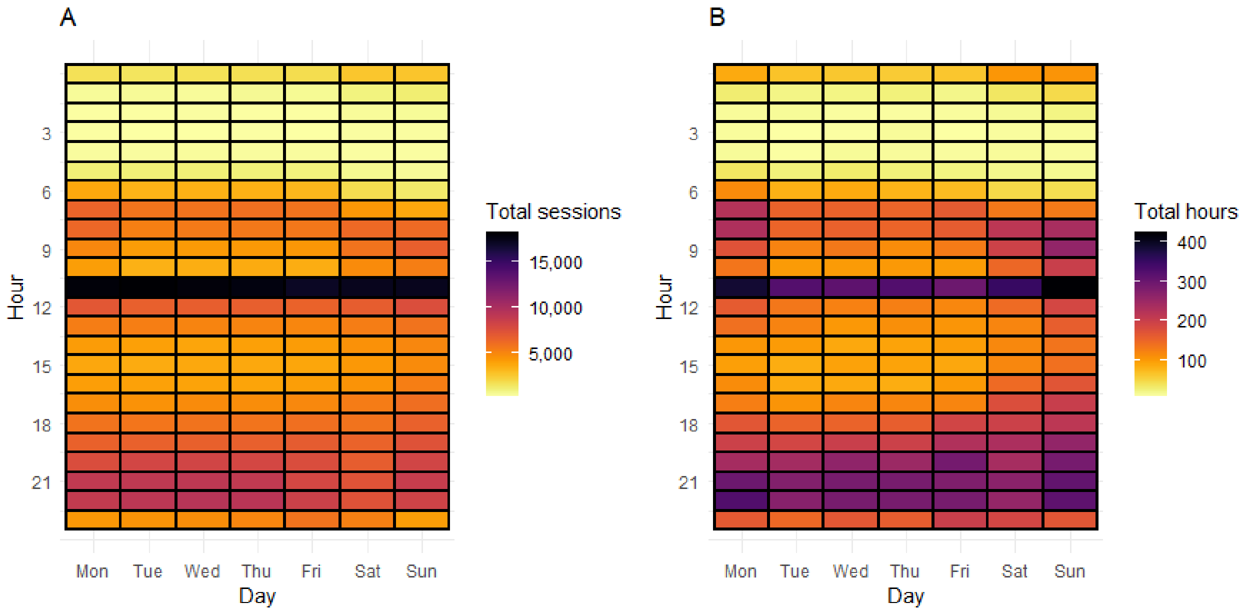
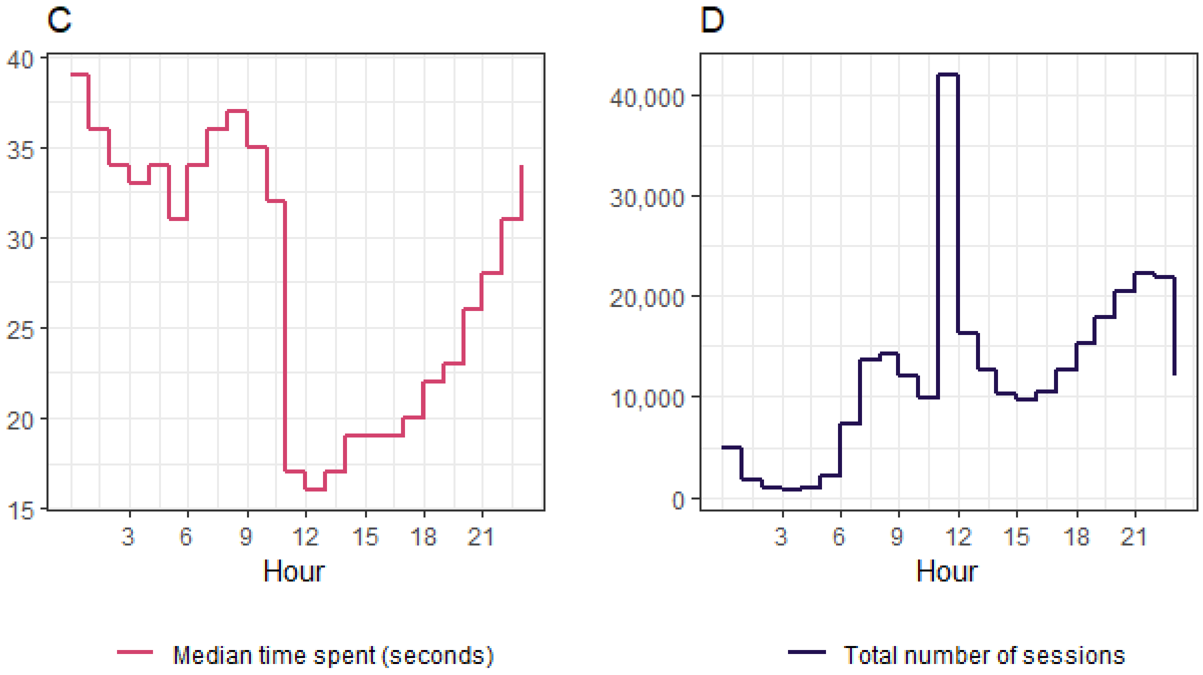
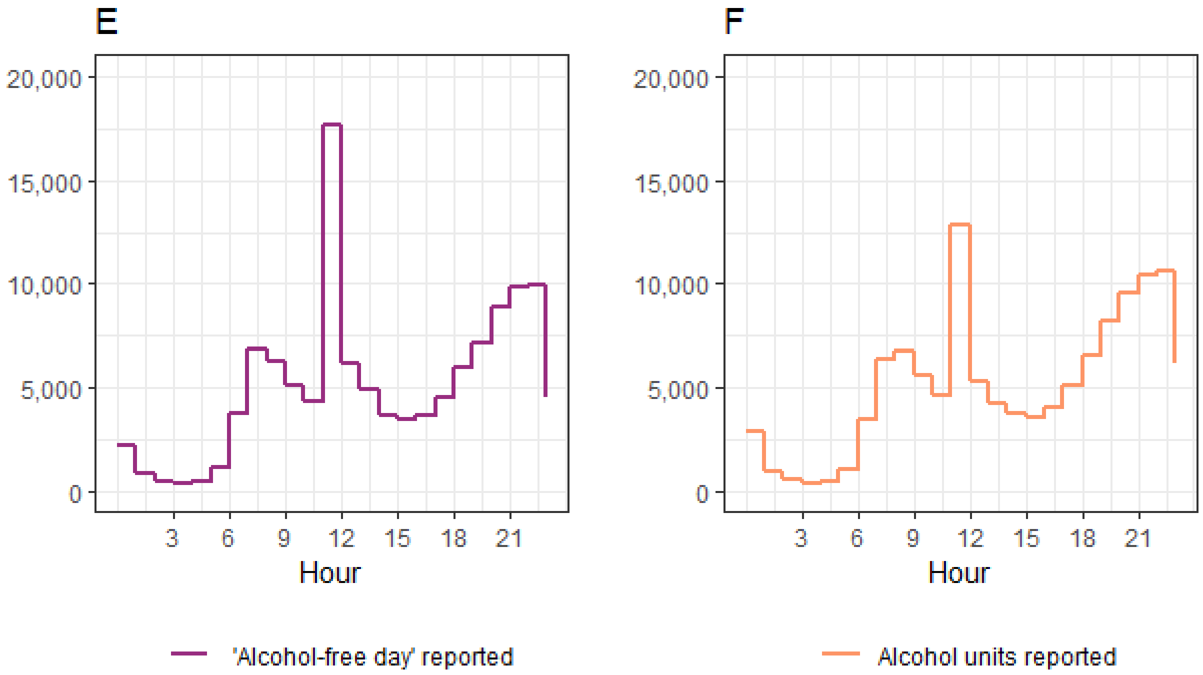


Figure 2. Median time spent on the *Drink Less* app per session and frequency distribution of sessions.



In Figure 3, plot E shows the frequency distribution of entering an alcohol-free day and plot F shows the frequency of entering a drink record. Timeline plots show the hour of the day on the x-axis. There are more alcohol-free days entered between 11 AM to 12 PM than drink records made, which suggests that the notification is more strongly associated with entering *alcohol-free days* than entering *alcohol units consumed*. Both outcomes see similar prominent, natural peaks in the evenings, with an additional smaller peak in the mornings.

Figure 3. Frequency distributions of when alcohol-free days and alcohol units are recorded during the day.



Visualization of Engagement Clusters

A total of 3 clusters emerged from the k-modes clustering. This was based on the measure *did the user open the app?* (binary, yes or no) for the first 30 days after download.

Figure 4 plots the probability of use of the app, stratified by cluster, over time (number of days after download). The 3 ribbons represent the probability of use of the app for each engagement cluster, with 95% CI. On the basis of the observed pattern of engagement, we named the 3 clusters as fast disengagers (10,903/19,233, 56.68%), slow disengagers (3679/19,233, 19.12%), and engagers (4651/19,233, 24.18%). The optimal number of clusters was determined by the elbow method and silhouette method (Multimedia Appendix 7). The silhouette method suggested that the optimal number of clusters was 2, whereas the elbow method suggested 3 clusters. Comparing the results under 2 and 3 clusters showed that the slow disengagers and engagers groups identified under 3 clusters were essentially a subdivision of 1 cluster in the 2-cluster model. We chose to retain 3 clusters based on observed differences in the trajectory of engagement over time between the 2 groups—the engagers and slow disengagers.

The probability of using the app 30 days after download for engagers was 0.69 (95% CI 0.67-0.70), slow disengagers was 0.10 (95% CI 0.10-0.11), and fast disengagers was 0.01 (95% CI 0.01-0.02).

Figure 4. Probability of use on day after download by cluster group with 95% CIs.

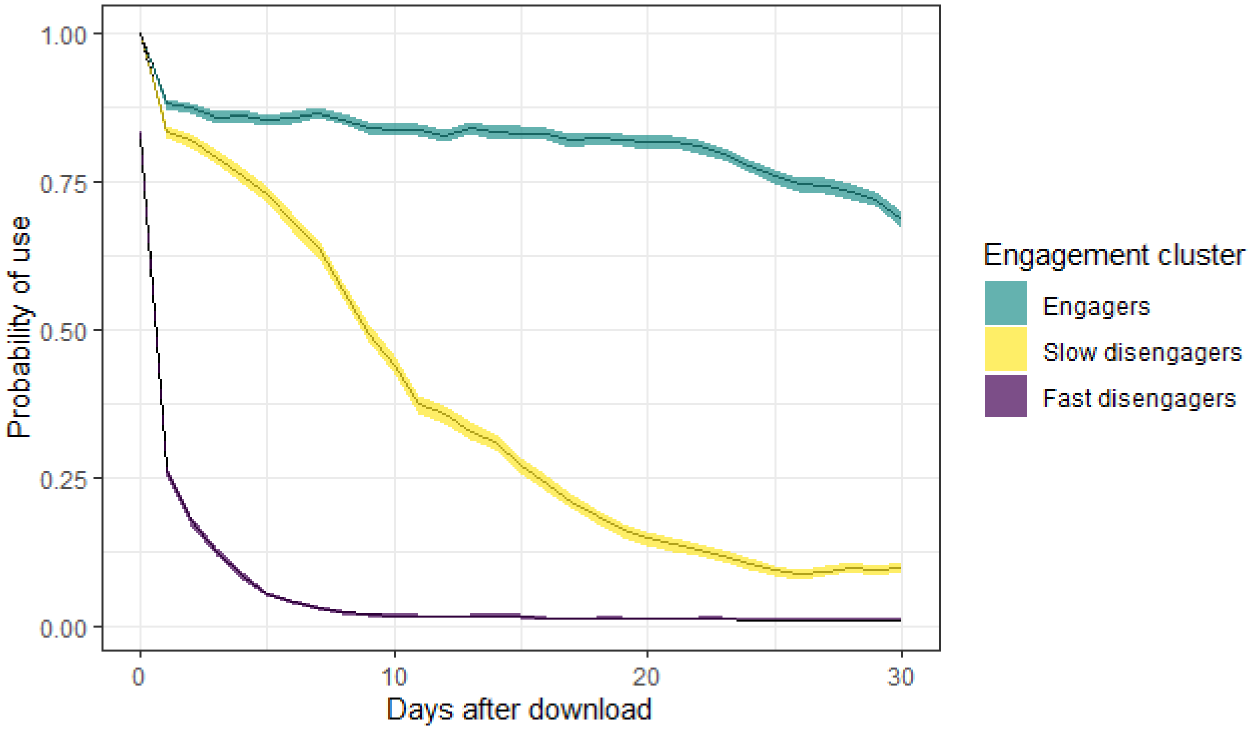


Table 2 shows the distribution of user characteristics across the engagement clusters. The median number of sessions was lowest for the fast disengagers and much higher for the engagers. Engagers are, on average, more likely to be older, male, working in nonmanual employment, and more likely to report a lower AUDIT risk zone, compared with users within the fast disengagers and slow disengagers clusters.

Table 2. User characteristics by cluster group.

User characteristics	Fast disengagers (n=10,903)	Slow disengagers (n=3679)	Engagers (n=4651)
Male, n (%)	4991 (45.78)	1920 (52.49)	2629 (56.53)
Age (years), mean (SD)	43.7 (11.57)	43.2 (10.91)	45.4 (10.57)
Employment type (nonmanual), n (%)	7567 (69.40)	2659 (72.28)	3566 (76.67)
AUDIT[a] risk zone, n (%)			
Hazardous (8-15)	5016 (46.01)	1577 (42.86)	2365 (50.85)
Harmful (16-19)	2155 (19.77)	798 (21.69)	996 (21.41)
At risk of alcohol dependence (20+)	3732 (34.23)	1304 (35.44)	1290 (27.74)
Number of sessions per user, median (25th-75th percentile)	3 (1-6)	18 (12-28)	88 (51-175)

[a] AUDIT: Alcohol Use Disorders Identification Test.

Table 3 provides the estimated associations between exposure to the notification and app use, based on Set B data. Over the first 30 days after day of download, the probability of using the app in the hour after the delivery of the notification was approximately 4 times higher than the probability of using the app in the hour before. All models of the estimated associations between exposure to the notification and app use are adjusted for the continuous variables of age, days after download, baseline AUDIT score, and the categorical variables of employment type and sex. The cluster-specific effects included an effect moderation of the exposure to the notification by cluster group, and the days after download effects included an effect moderation of the exposure by days after download. The adjusted estimated risk ratio was 4.21 (95% CI 4.07-4.36), and the estimated risk ratio was higher among engagers (Wald test *P* value: fast disengagers vs engagers *P*=.001 slow disengagers vs engagers *P*<.001, slow disengagers vs fast disengagers *P*=.44).

Table 4 shows the estimated association between exposure to the notification and opening of the app in the 3 clusters at 3 different time points (days 1, 7, and 30).

Table 3. Estimated associations between exposure to the notification and app use.

Model	Exposure to notification, estimated relative risk ratio (95% CI)
Unadjusted model	4.22 (4.13-4.31)
Adjusted model[a]	4.21 (4.07-4.36)
Days after download[b]	
Day 1	3.93 (3.77-4.10)
Day 7	4.07 (3.93-4.22)
Day 30	4.67 (4.38-4.98)
Cluster[c]	
Fast disengagers	3.97 (3.70-4.25)
Slow disengagers	3.82 (3.60-4.03)
Engagers	4.38 (4.18-4.59)

[a] Adjusted for days after download, employment type, sex, age, and baseline Alcohol Use Disorders Identification Test (AUDIT) score.

[b] Estimated from the model including the interaction effect of exposure to the notification by days after download, adjusted for employment type, sex, age, and baseline AUDIT score.

[c] Estimated from the model including the interaction effect of exposure to the notification by cluster, adjusted for days after download, employment type, sex, age, and baseline AUDIT score.

Table 4. Estimated risk ratio with 95% CI for the associations between exposure to the notification and app use within each cluster, at 3 time points (days 1, 7, and 30).

Clusters	Risk ratio at day 1 (95% CI[a])	Risk ratio at day 7 (95% CI[a])	Risk ratio at day 30 (95% CI[a])
Fast disengagers	3.66 (3.33-4.02)	3.83 (3.57-4.11)	4.58 (3.86-5.43)
Slow disengagers	4.18 (3.85-4.54)	3.87 (3.64-4.12)	2.89 (2.43-3.43)
Engagers	4.05 (3.82-4.30)	4.22 (4.01-4.43)	4.90 (4.56-5.26)

[a] Interaction effect of exposure to the notification and days after download, an interaction effect of exposure to the notification and cluster, and a three-way interaction effect of exposure to the notification, cluster, and days after download, adjusted for employment type, sex, age, and baseline Alcohol Use Disorders Identification Test score.

Visualization of Time to Disengagement

Kaplan-Meier plots, both overall and stratified by clusters, were plotted to show days to disengagement, defined as 7 or more consecutive days of no use, for the first 365 days after downloading *Drink Less*.

In Figures 5 and 6, the x-axis depicts the number of days after download, ranging from 0 to 365, and the y-axis depicts the survival probability, which is the proportion of users who have not disengaged. The dashed lines at the 0.5 survival probability mark shows the time (days) up to when 50%of each cluster has disengaged. Each hash in the plot represents a right-censored user. The number at risk represents the users in the clusters who remain engaged over the year. In Figure 5, we see that 50.00% (9617/19,233) of users have disengaged at 22 days from download, and Figure 6 shows the divergence of longer-term engagement between clusters. The median number of days to disengagement for engagers was 132 days (95% CI 128-137), slow disengagers was 26 days (95% CI 24-29), and fast disengagers was 3 days (95% CI 2-3).

Figure 5. Time to disengagement (defined as the first day of 7 or more consecutive days of no use) for all users.

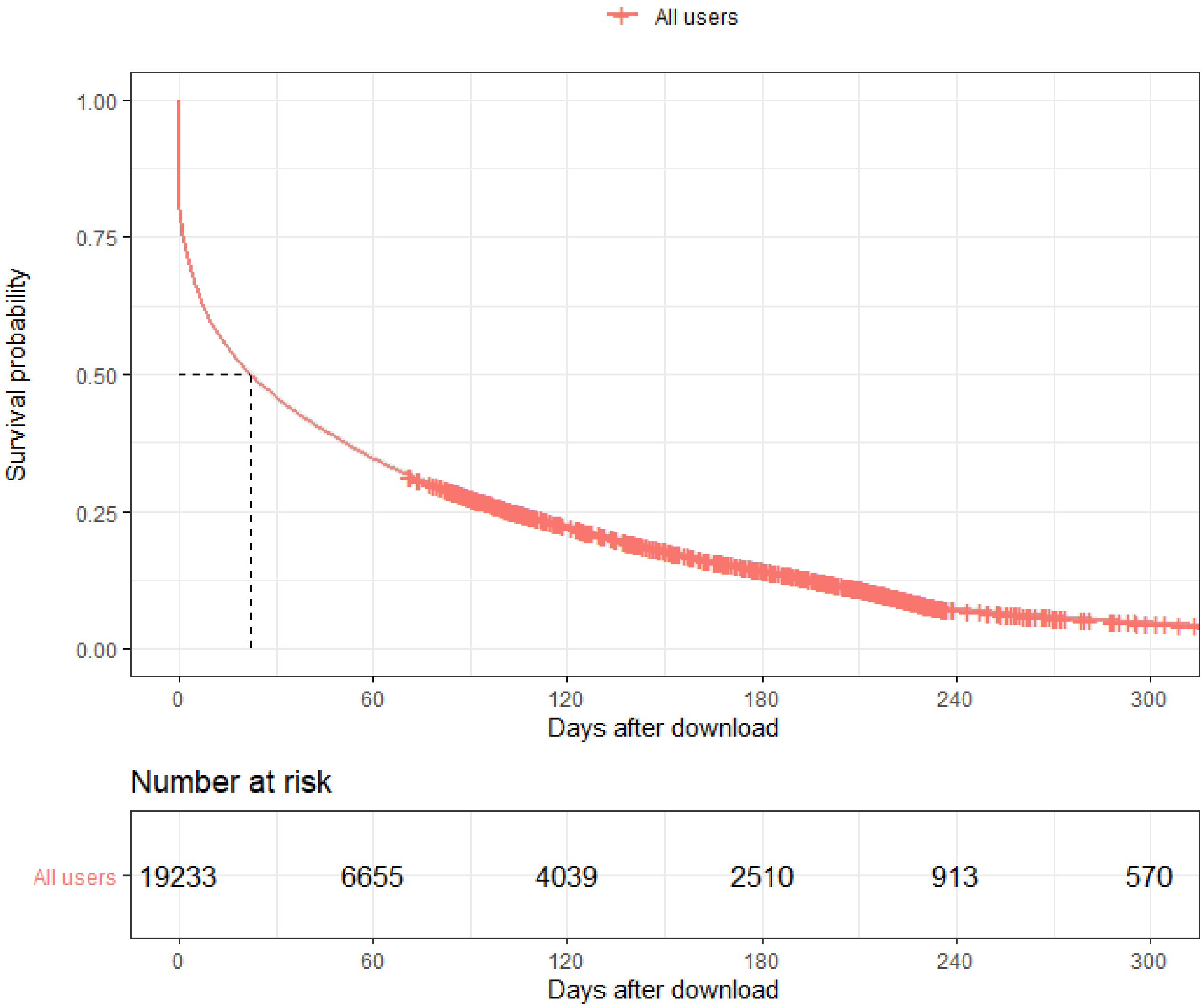
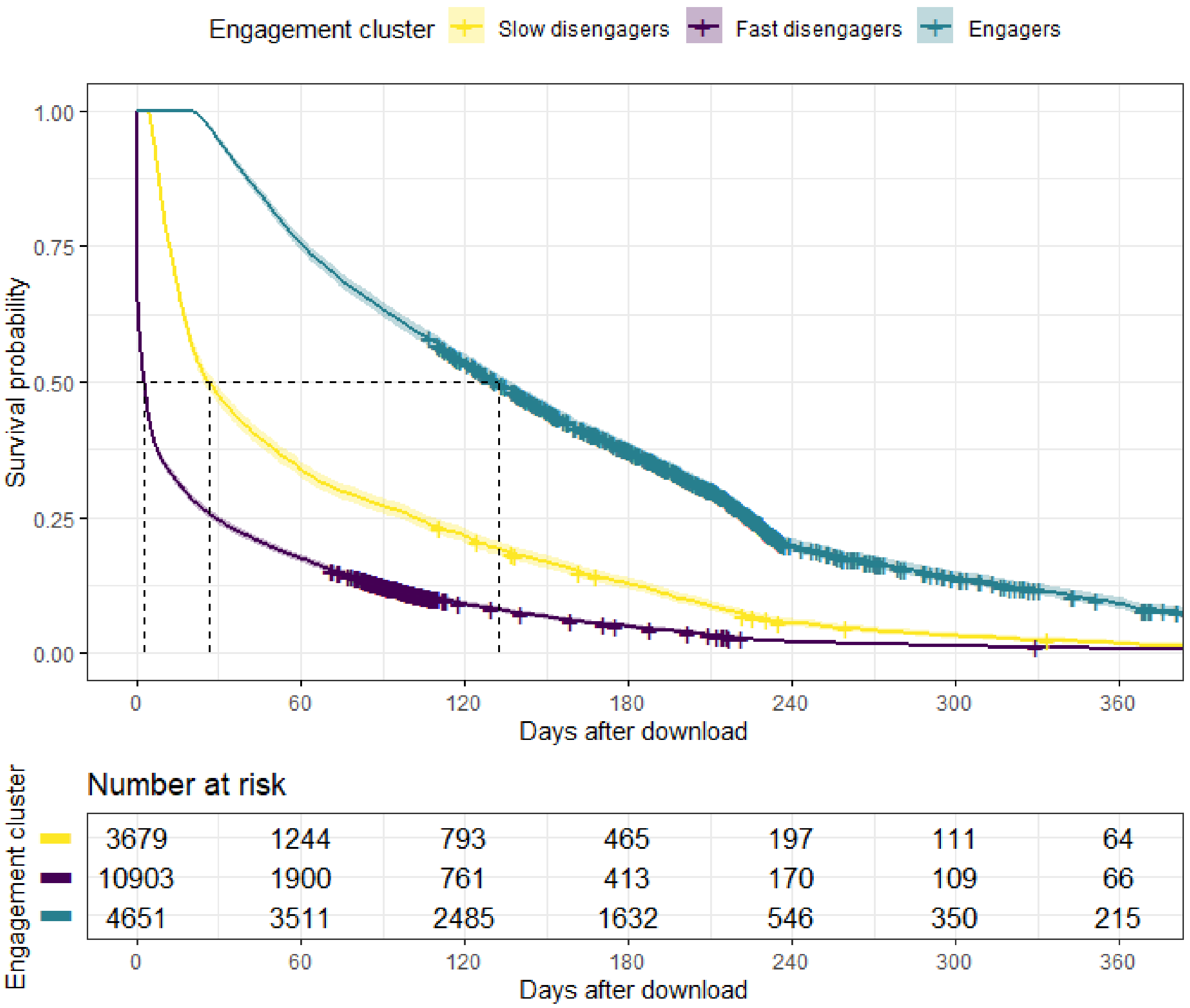


Figure 6. Time to disengagement (defined as the first day of 7 or more consecutive days of no use) by the engagement cluster.



Engagement cluster	0	60	120	180	240	300	360
🟨	3679	1244	793	465	197	111	64
🟪	10903	1900	761	413	170	109	66
🟦	4651	3511	2485	1632	546	350	215

Discussion

Principal Findings

Visualizations provided important insights into how users engage with the behavior change app *Drink Less*. They revealed a strong association between delivery of the daily push notification (sent at 11 AM) and use in the next hour, suggesting that the push notification strongly influences how users engage with *Drink Less* in the immediate hour after the notification is sent. Push notifications (sometimes known as ecological momentary interventions) are programmed messages sent to a user by the app and are commonly employed within behavior change apps to both monitor and provide support to people at risk of harmful alcohol consumption [3,10,59]. Push notifications are a time-varying component of *Drink Less* that can be further optimized to become just-in-time adaptive interventions that rely on decision rules in the provision of real-time support and can learn and adapt to the contextual and psychological circumstances of individuals over time [59]. Previous research has found that notifications are important components that influence engagement with behavior change apps [60-62]. This includes an ecological momentary assessment study with *Drink Less*, which found that establishing a daily routine is important for maintaining engagement and that the daily push notification supports such routines [63]. This study also found that time-varying, endogenous factors of motivation and perceived usefulness of the app were the most consistent predictors of engagement.

Our analysis suggested an approximate adjusted four-fold increase in the probability of using the app in the immediate hour following the notification (11 AM to noon) compared with the preceding hour (10 AM to 11 AM). For the 1 in 5 users belonging to the slow disengagers group, two interesting findings emerged. Firstly, the association between the notification and opening *Drink Less* in the subsequent hour decreased over time, and secondly, patterns of engagement for this group show, on average, a high probability of use over the first week but low probability after the second week. A possible reason for this decline in probability of use and association of the notification and use could be habituation to the daily notification, or turning the notification off. Importantly, we hypothesize that optimizing the notification policy may generate higher rates of engagement for this group.

Future Research to Understand and Optimize the Notification Policy

To carefully create decision rules for the policy to evolve from an ecological momentary intervention to a just-in-time adaptive intervention, we will undertake a micro-randomized trial (MRT). The aim of the MRT is to further develop the push notification policy to improve engagement by targeting internal or external contextual circumstances that either influence excessive drinking (states of vulnerability) or events of engagement with the app (states of acceptability and opportunity) [64]. Visualization of engagement data helped inform the design of our MRT.

Table 5 summarizes how the visualizations from this exploratory research informed the design of our forthcoming MRT.

Primarily, this research guided our decision to shift the delivery time from 11 AM to 8 PM to exploit the potential increase in vulnerability to excess drinking, in an opportune and acceptable moment to engage with *Drink Less* [64]. To avoid the risk of an underpowered MRT, the expected effect size used in the sample size calculation of our MRT is based on a more conservative model, with the control defined as use between 9 AM and 11 AM and the treatment defined as use between 11 AM and 1 PM. This means that the MRT is powered to detect a marginal effect, quantified as a risk ratio, of sending a notification (compared with sending no notification) of 2.16 on user engagement rather than 4.22. We added 2 parallel arms to the MRT to provide an assessment of how engagement with *Drink Less* evolves over time when no notifications are provided and an exchangeable sample to compare the current policy of delivering a fixed notification daily, to a random notification policy, varying the content and sequence of notifications.

Table 5. Linking visualization to the design of a micro-randomized trial.

What we learnt from these analyses	Which visualization or analyses showed us this	How this informed the design of our randomized trial
The present notification appears to be a key driver of engagement	• Figure 1, plot A: heat map of total sessions. • Figure 1, plot B: heat map of total time on app (hours) • Table 4: estimated risk ratio with 95% CI for the associations between exposure to the notification and app use within each cluster, at 3 time points (days 1, 7, and 30)	We chose to undertake a micro-randomized trial to both understand the causal effect of the notification on engagement, and to further optimize the delivery of notifications with respect to time-varying covariates, notifications, and outcomes
The impact of the notification seems to be strongest in the hour preceding delivery	Figure 1, plot A: heat map of total sessions	We set the time window to measure the proximal (ie, near-term) effect as 1 hour after delivery
Evenings seem to be an opportune and acceptable moment to engage with *Drink Less.* It is also a time of increased vulnerability to excess drinking	Figure 1, plot B: heat map of total time on app (hours)	We moved the delivery time of the notification to 8 PM
The notification may encourage the reporting of *alcohol-free days* more than *drink consumed*. This may be due to competing pressures for time at 11 AM	Figure 3: frequency distributions of when alcohol-free days and alcohol units are recorded	We intervened in the evenings to see if this is a more acceptable and opportune time to report drinks consumed
The notification may reduce the median time per session during the reminder of the day	Figure 2: line plot of median time spent on app (seconds)	We included a *no-notification* arm in our trial to capture a momentary assessment of engagement when no notifications are sent
The depth of engagement with *Drink Less* is low	Multimedia Appendix 2: summaries of use by module for all users	We trialed new notifications which target the perceived usefulness of *Drink Less* to encourage broader engagement
Slow disengagers (3679/19,233, 19.13%) have a high probability of engagement during the first week, but by day 30, this group has a low probability, suggesting a loss of motivation	Figure 4: probability of use on day after download by cluster group	We tested 30 new messages to increase novelty and motivation to remain engaged with *Drink Less* (Multimedia Appendix 8)
Exogenous impacts, such as public health campaigns, are likely to influence the cohort of users over time	Figure 6: time to disengagement (defined as the first day of 7 or more consecutive days of no use) by the engagement cluster	We included a *standard app version* arm in the trial, to provide an exchangeable sample to compare the fixed and random notification policies

Limitations

This paper details exploratory research. Our estimates of the association between the notification and opening *Drink Less* do not represent a causal effect on engagement, as we are unable to account for systematic differences in use between the 2 periods that are unrelated to the notification. A randomized trial will allow for the causal effect of the notification to be understood. We also found that simple, accessible visualizations achieved our goal of understanding important patterns of engagement; however, when managing denser streams of data, more complex visualizations may be required.

An additional limitation is that disengagement is defined as a period of no use for 7 or more consecutive days and is considered as a one-time event instead of a repeated event; hence, the Kaplan-Meier plots are interpreted for the survival event *disengagement for the first time*. However, a proportion of users repeatedly disengage and then re-engage with *Drink Less*. It is not uncommon that even after disengaging a number of times, users re-engage for long, continuous spells of use with *Drink Less*. We aim to explore this in future research by visualizing the nature of repeated reengagement with accessible graphical applications and available shared toolsets [34]. An additional limitation is that we did not track whether users subsequently turned off or altered the delivery time of their notifications.

Conclusions

Identifying patterns of engagement from voluminous, temporally dense data presents various challenges for researchers and practitioners. The summarization of such data with heat maps, timeline plots, and Kaplan-Meier plots can provide a clear picture of daily, weekly, and long-term patterns of use over time with a behavior change app. Optimizing engagement is a priority for many behavior change apps, and these visualizations provide a way to identify the key features of how this version of a behavior change app is engaged with.

For *Drink Less*, we have demonstrated the important role of visualizations by showing how these clearly identified how behavioral engagement varies over the day of the week and hour of the day, along with when users first disengage. The visualizations revealed that the daily notification is likely to strongly influence engagement with *Drink Less*. Both the average probability of use over 30 days and the association between use and the notification remained high for users in the engagers cluster yet steadily declined over time for users in the slow disengagers cluster. This suggests that a fixed notification policy can be effective for maintaining engagement for some users but ineffective for others. It is now our priority to understand the causal effect of the notification on engagement and to consider further optimizing the push notification policy to contextual circumstances of individuals over time to inform the development of a just-in-time adaptive intervention. The MRT aims to inform the development of decision rules to tailor the notification policy to individuals over time, with details found in our protocol [65].

Acknowledgments

The development of the *Drink Less* app was funded by the National Institute for Health Research (NIHR) School for Public Health Research, Society for the Study of Addiction, Cancer Research UK, and the UK Centre for Tobacco and Alcohol Studies (UKCTAS). The NIHR School for Public Health Research is a partnership between the Universities of Sheffield, Bristol, Cambridge, and Imperial; University College London; The London School for Hygiene and Tropical Medicine; LiLaC: a collaboration between the Universities of Liverpool and Lancaster; and Fuse—The Centre for Translational Research in Public Health: a collaboration between Newcastle, Durham, Northumbria, Sunderland, and Teesside Universities. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. UKCTAS is part of UK Clinical Research Collaboration, a Public Health Research Centre of Excellence. Funding from the Medical Research Council, British Heart Foundation, Cancer Research UK, Economic and Social Research Council, and the NIHR under the auspices of the UK Clinical Research Collaboration is gratefully acknowledged.

The authors would like to thank Dr Dave Crane for his important role in the development and factorial screening trial of the *Drink Less* app and the National University of Singapore's Institute of Mathematical Sci*ences* for funding LB and TQ's visit to the program on Statistical Methods for Developing Personalized Mobile Health Interventions. The authors would also like to thank the UCL Tobacco and Alcohol Research Group for their helpful feedback on the draft of this paper.

LB is supported by a PhD studentship funded by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2-R18).

CG is funded by the NIHR, the NIHR School for Public Health Research and Cancer Research UK (CRUK: C1417/A22962).

TQ is supported by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award number R01AA23187; National Institute on Drug Abuse of the National Institutes of Health under award number P50DA039838; National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number U54EB020404; and National Cancer Institute of the National Institutes of Health under award number U01CA229437. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

OP is funded by Cancer Research UK (CRUK: C1417/A22962).

EW is supported by the MRC (MR/S01442X/1) and Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), and British Heart Foundation and Wellcome. HP receives funding from Public Health England, the NIHR and NHS England.

Conflicts of Interest

CG is involved in the development and testing of the *Drink Less* app, and OP is an unpaid member of the scientific committee for the Smoke Free app. HP has received consultancy fees from Babylon Health. All other authors declare no conflict of interests.

Multimedia Appendix 1
Privacy notice.
[DOCX File , 14 KB - jmir_v22i12e23369_app1.docx]

Multimedia Appendix 2
Summaries of Use by Module for all users (n=19,233).
[DOCX File , 13 KB - jmir_v22i12e23369_app2.docx]

Multimedia Appendix 3
3D rotating heatmap of when sessions begin.
[MP4 File (MP4 Video), 2999 KB - jmir_v22i12e23369_app3.mp4]

Multimedia Appendix 4
3D rotating graph total time on app per session.
[MP4 File (MP4 Video), 3167 KB - jmir_v22i12e23369_app4.mp4]

Multimedia Appendix 5
Plot D animated over time.
[MP4 File (MP4 Video), 3436 KB - jmir_v22i12e23369_app5.mp4]

Multimedia Appendix 6
Heat map of when downloads occur.
[DOCX File , 29 KB - jmir_v22i12e23369_app6.docx]

Multimedia Appendix 7
Plots of the elbow method and silhouette method.
[DOCX File , 38 KB - jmir_v22i12e23369_app7.docx]

Multimedia Appendix 8
Notification content of new message bank to be trialled in the micro-randomized trial.
[DOCX File , 14 KB - jmir_v22i12e23369_app8.docx]

References

1. UK Chief Medical Officers' Low Risk Drinking Guidelines. Department of Health. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/545937/UK_CMOs__report.pdf [accessed 2020-06-09]
2. Waters AJ, Schoenmakers TM, Snelleman M, Szeto EH, Franken IH, Hendriks VM, et al. Affect, motivation, temptation, and drinking among alcohol-dependent outpatients trying to maintain abstinence: An Ecological Momentary Assessment study. Drug Alcohol Depend 2020 Jan 1;206:107626. [doi: 10.1016/j.drugalcdep.2019.107626] [Medline: 31786398]
3. Beckjord E, Shiffman S. Background for real-time monitoring and intervention related to alcohol use. Alcohol Res 2014;36(1):9-18 [FREE Full text] [Medline: 26258996]
4. Dvorak RD, Pearson MR, Day AM. Ecological momentary assessment of acute alcohol use disorder symptoms: associations with mood, motives, and use on planned drinking days. Exp Clin Psychopharmacol 2014 Aug;22(4):285-297 [FREE Full text] [doi: 10.1037/a0037157] [Medline: 24932896]
5. Cook MA, Newins AR, Dvorak RD, Stevenson BL. What about this time? Within- and between-person associations between drinking motives and alcohol outcomes. Exp Clin Psychopharmacol 2020 Oct;28(5):567-575. [doi: 10.1037/pha0000332] [Medline: 31670547]
6. Chakravorty S, Kuna ST, Zaharakis N, O'Brien CP, Kampman KM, Oslin D. Covariates of craving in actively drinking alcoholics. Am J Addict 2010;19(5):450-457 [FREE Full text] [doi: 10.1111/j.1521-0391.2010.00067.x] [Medline: 20716308]
7. Lac A, Handren L, Crano WD. Conceptualizing and measuring weekend versus weekday alcohol use: item response theory and confirmatory factor analysis. Prev Sci 2016 Oct;17(7):872-881 [FREE Full text] [doi: 10.1007/s11121-016-0685-9] [Medline: 27488456]

8. Lac A, Luk JW. Pathways from positive, negative, and specific alcohol expectancies to weekday and weekend drinking to alcohol problems. Prev Sci 2019 Jul;20(5):800-809 [FREE Full text] [doi: 10.1007/s11121-019-0986-x] [Medline: 30685802]
9. Cohn A, Hunter-Reel D, Hagman BT, Mitchell J. Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. Alcohol Clin Exp Res 2011 Dec;35(12):2209-2215 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01571.x] [Medline: 21689119]
10. Morgenstern J, Kuerbis A, Muench F. Ecological momentary assessment and alcohol use disorder treatment. Alcohol Res 2014;36(1):101-109 [FREE Full text] [Medline: 26259004]
11. Petry N, Meredith S, Alessi S. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. Adv Health Care Technol 2015 Dec:47. [doi: 10.2147/ahct.s65791]
12. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42-47 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
13. Freisthler B, Lipperman-Kreda S, Bersamin M, Gruenewald PJ. Tracking the when, where, and with whom of alcohol use: integrating ecological momentary assessment and geospatial data to examine risk for alcohol-related problems. Alcohol Res 2014;36(1):29-38 [FREE Full text] [Medline: 26258998]
14. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
15. Forbes G, Newton S, Cantalapiedra Calvete C, Birch J, Dodds J, Steed L, et al. MEMPHIS: a smartphone app using psychological approaches for women with chronic pelvic pain presenting to gynaecology clinics: a randomised feasibility trial. BMJ Open 2020 Mar 12;10(3):e030164 [FREE Full text] [doi: 10.1136/bmjopen-2019-030164] [Medline: 32165549]
16. Torous J, Michalak EE, O'Brien HL. Digital health and engagement-looking behind the measures and methods. JAMA Netw Open 2020 Jul 1;3(7):e2010918 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]
17. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun 13;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
18. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 7;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
19. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mhealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
20. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
21. Collins LM, Dziak JJ, Li R. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. Psychol Methods 2009 Sep;14(3):202-224 [FREE Full text] [doi: 10.1037/a0015826] [Medline: 19719358]
22. Collins LM, Chakraborty B, Murphy SA, Strecher V. Comparison of a phased experimental approach and a single randomized clinical trial for developing multicomponent behavioral interventions. Clin Trials 2009 Feb;6(1):5-15 [FREE Full text] [doi: 10.1177/1740774508100973] [Medline: 19254929]
23. Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, Smith SS, et al. The multiphase optimization strategy for engineering effective tobacco use interventions. Ann Behav Med 2011 Apr;41(2):208-226 [FREE Full text] [doi: 10.1007/s12160-010-9253-x] [Medline: 21132416]
24. Anderson R. New MRC guidance on evaluating complex interventions. Br Med J 2008 Oct 22;337(oct22 1):a1937. [doi: 10.1136/bmj.a1937]
25. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. Br Med J 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
26. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
27. Garnett C, Crane D, West R, Brown J, Michie S. The development of Drink Less: an alcohol reduction smartphone app for excessive drinkers. Transl Behav Med 2019 Mar 1;9(2):296-307 [FREE Full text] [doi: 10.1093/tbm/iby043] [Medline: 29733406]
28. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]

29. Ola O, Sedig K. The challenge of big data in public health: an opportunity for visual analytics. Online J Public Health Inform 2014;5(3):223 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4933] [Medline: 24678376]
30. Chui KK, Wenger JB, Cohen SA, Naumova EN. Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs. PLoS One 2011 Feb 15;6(2):e14683 [FREE Full text] [doi: 10.1371/journal.pone.0014683] [Medline: 21347221]
31. Ola O, Sedig K. Beyond simple charts: design of visualizations for big health data. Online J Public Health Inform 2016;8(3):e195 [FREE Full text] [doi: 10.5210/ojphi.v8i3.7100] [Medline: 28210416]
32. Gelman A, Pasarica C, Dodhia R. Let's practice what we preach. Am Stat 2002 May;56(2):121-130. [doi: 10.1198/000313002317572790]
33. Ballarini NM, Chiu Y, König F, Posch M, Jaki T. A critical review of graphics for subgroup analyses in clinical trials. Pharm Stat 2020 Sep;19(5):541-560. [doi: 10.1002/pst.2012] [Medline: 32216035]
34. Morrison C, Doherty G. Analyzing engagement in a web-based intervention platform through visualizing log-data. J Med Internet Res 2014 Nov 13;16(11):e252 [FREE Full text] [doi: 10.2196/jmir.3575] [Medline: 25406097]
35. Arden-Close EJ, Smith E, Bradbury K, Morrison L, Dennison L, Michaelides D, et al. A visualization tool to analyse usage of web-based interventions: the example of positive online weight reduction (power). JMIR Hum Factors 2015 May 19;2(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.4310] [Medline: 27026372]
36. Sharmin M, Raij A, Epstien D, Nahum-Shani I, Beck JG, Vhaduri S, et al. Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. Proc ACM Int Conf Ubiquitous Comput 2015 Sep;2015:505-516 [FREE Full text] [doi: 10.1145/2750858.2807537] [Medline: 26539566]
37. Chen AT, Wu S, Tomasino KN, Lattie EG, Mohr DC. A multi-faceted approach to characterizing user behavior and experience in a digital mental health intervention. J Biomed Inform 2019 Jun;94:103187 [FREE Full text] [doi: 10.1016/j.jbi.2019.103187] [Medline: 31026595]
38. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 1;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
39. Bohn MJ, Babor TF, Kranzler HR. The alcohol use disorders identification test (AUDIT): validation of a screening instrument for use in medical settings. J Stud Alcohol 1995 Jul;56(4):423-432. [doi: 10.15288/jsa.1995.56.423] [Medline: 7674678]
40. How a Web Session is Defined in Analytics. Google Analytics. 2017. URL: https://support.google.com/analytics/answer/2731565 [accessed 2020-04-01]
41. Du H, Venkatakrishnan A, Youngblood GM, Ram A, Pirolli P. A group-based mobile application to increase adherence in exercise and nutrition programs: a factorial design feasibility study. JMIR Mhealth Uhealth 2016 Jan 15;4(1):e4 [FREE Full text] [doi: 10.2196/mhealth.4900] [Medline: 26772910]
42. R: A Language and Environment for Statistical Computing. R Core Team. 2020. URL: https://www.R-project.org/ [accessed 2020-11-25]
43. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
44. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, USA: Springer-Verlag; 2016.
45. Morgan-Wall T. Create Maps and Visualize Data in 2D and 3D. CRAN. 2020. URL: https://CRAN.R-project.org/package=rayshader [accessed 2020-04-01]
46. Garnier S. Default Color Maps From 'matplotlib'. CRAN. 2018. URL: https://CRAN.R-project.org/package=viridis, [accessed 2020-04-01]
47. Weihs C, Ligges U, Karsten L, Nils R. klaR analyzing German business cycles. In: Baier D, Decker R, Schmidt-Thieme L, editors. Data Analysis and Decision Support. Berlin, Germany: Springer-Verlag; 2005:335-343.
48. Kassambara A, Kosinski M, Przemyslaw B. Drawing Survival Curves Using 'ggplot2'. CRAN. https://CRAN.R-project.org/package=survminer; 2020. URL: https://CRAN.R-project.org/package=survminer, [accessed 2020-04-01]
49. Lin Pedersen T, Robinson D. A Grammar of Animated Graphics. CRAN. 2020. URL: https://CRAN.R-project.org/package=gganimate [accessed 2020-11-25]
50. Lin Pederson T. The Composer of Plots. CRAN. 2019. URL: https://CRAN.R-project.org/package=patchwork [accessed 2020-04-01]
51. Huang Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining Knowl Discovery 1998;2(3):283-304. [doi: 10.1023/A:1009769707641]
52. Bishop CM. In: Jordan M, Kleinberg J, Scholkopf B, editors. Pattern Recognition and Machine Learning. New York, USA: Springer; 2006.
53. Yock AD, Kim G. Technical Note: Using k-means clustering to determine the number and position of isocenters in MLC-based multiple target intracranial radiosurgery. J Appl Clin Med Phys 2017 Sep;18(5):351-357 [FREE Full text] [doi: 10.1002/acm2.12139] [Medline: 28727284]
54. Handl J, Knowles J, Kell DB. Computational cluster validation in post-genomic data analysis. Bioinformatics 2005 Aug 1;21(15):3201-3212. [doi: 10.1093/bioinformatics/bti517] [Medline: 15914541]

55. Lengyel A, Botta-Dukát Z. Silhouette width using generalized mean-a flexible method for assessing clustering efficiency. Ecol Evol 2019 Dec;9(23):13231-13243 [FREE Full text] [doi: 10.1002/ece3.5774] [Medline: 31871641]
56. Morris TP, Jarvis CI, Cragg W, Phillips PP, Choodari-Oskooei B, Sydes MR. Proposals on Kaplan-Meier plots in medical research and a survey of stakeholder views: KMunicate. BMJ Open 2019 Sep 30;9(9):e030215 [FREE Full text] [doi: 10.1136/bmjopen-2019-030215] [Medline: 31575572]
57. Royston P, Parmar MK, Altman DG. Visualizing length of survival in time-to-event studies: a complement to Kaplan-Meier plots. J Natl Cancer Inst 2008 Jan 16;100(2):92-97. [doi: 10.1093/jnci/djm265] [Medline: 18182621]
58. LIANG K, ZEGER SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
59. Bidargaddi N, Schrader G, Klasnja P, Licinio J, Murphy S. Designing m-Health interventions for precision mental health support. Transl Psychiatry 2020 Jul 7;10(1):222 [FREE Full text] [doi: 10.1038/s41398-020-00895-2] [Medline: 32636358]
60. Morrison LG, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS ONE 2017 Jan 3;12(1):e0169162. [doi: 10.1371/journal.pone.0169162]
61. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
62. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
63. Perski O, Naughton F, Garnett C, Blandford A, Beard E, West R, et al. Do daily fluctuations in psychological and app-related variables predict engagement with an alcohol reduction app? A series of n-of-1 studies. JMIR Mhealth Uhealth 2019 Oct 2;7(10):e14098 [FREE Full text] [doi: 10.2196/14098] [Medline: 31579022]
64. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
65. Bell L, Garnett C, Qian T, Perski O, Potts HW, Williamson E. Notifications to improve engagement with an alcohol reduction app: protocol for a micro-randomized trial. JMIR Res Protoc 2020 Aug 7;9(8):e18690 [FREE Full text] [doi: 10.2196/18690] [Medline: 32763878]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
MRT: micro-randomized trial
NIHR: National Institute for Health Research
UKCTAS: UK Centre for Tobacco and Alcohol Studies

Edited by G Eysenbach; submitted 10.08.20; peer-reviewed by C Morrison, D Leightley, R Krukowski; comments to author 26.08.20; revised version received 08.10.20; accepted 28.10.20; published 11.12.20.

Please cite as:
Bell L, Garnett C, Qian T, Perski O, Williamson E, Potts HWW
Engagement With a Behavior Change App for Alcohol Reduction: Data Visualization for Longitudinal Observational Study
J Med Internet Res 2020;22(12):e23369
URL: http://www.jmir.org/2020/12/e23369/
doi:10.2196/23369
PMID:33306026

Original Paper

Acceptance of a Virtual Reality Headset Designed for Fall Prevention in Older Adults: Questionnaire Study

Nicolas Mascret[1], PhD; Lisa Delbes[1], MSc; Amélie Voron[1], MSc; Jean-Jacques Temprado[1], PhD; Gilles Montagne[1], PhD

[1] Aix Marseille Univ, CNRS, ISM, Marseille, France

Corresponding Author:
Nicolas Mascret, PhD
Aix Marseille Univ, CNRS, ISM
163 avenue de Luminy, case 910
Marseille, 13288 cedex 9
France
Phone: 33 618022983
Email: nicolas.mascret@univ-amu.fr

Abstract

Background: Falls are a common phenomenon among people aged 65 and older and affect older adults' health, quality of life, and autonomy. Technology-based intervention programs are designed to prevent the occurrence of falls and their effectiveness often surpasses that of more conventional programs. However, to be effective, these programs must first be accepted by seniors.

Objective: Based on the technology acceptance model, this study aimed to examine the acceptance among older adults before a first use of a virtual reality headset (VRH) used in an intervention program designed to prevent falls.

Methods: A sample of 271 French older adults (mean age 73.69 years, SD 6.37 years) voluntarily and anonymously filled out a questionnaire containing the focal constructs (perceived usefulness, perceived enjoyment, perceived ease of use, intention to use, fall-related self-efficacy, and self-avoidance goals) adapted to the VRH, which was designed to prevent falls.

Results: The results of the structural equation modeling analysis showed that intention to use the VRH was positively predicted by perceived usefulness, perceived enjoyment, and perceived ease of use. Perceived usefulness of the VRH was also negatively predicted by fall-related self-efficacy (ie, the perceived level of confidence of an individual when performing daily activities without falling) and positively predicted by self-avoidance goals (ie, participating in a physical activity to avoid physical regression).

Conclusions: A better understanding of the initial acceptance among older adults of this VRH is the first step to involving older adults in intervention programs designed to prevent falls using this kind of device.

(J Med Internet Res 2020;22(12):e20691) doi:10.2196/20691

KEYWORDS

technology acceptance model; acceptability; acceptance; virtual reality; elderly; fall; eHealth; self-efficacy; achievement goals

Introduction

Background

The number and proportion of people aged 65 years and older in the general population is likely to increase in the coming decades [1]. Among the numerous factors that may influence the quality of life and autonomy of older adults, falls are one of the most frequent and dramatic. Indeed, approximately 40% of people older than 65 years fall each year, about 2.5% of them will be hospitalized, and only one-half of those hospitalized will survive 1 year [2]. Consequently, research on fall prevention is of great scientific, practical, and socioeconomic interest, with an aim to help older adults to live longer independently as well as to reduce the burden on the health care system [3].

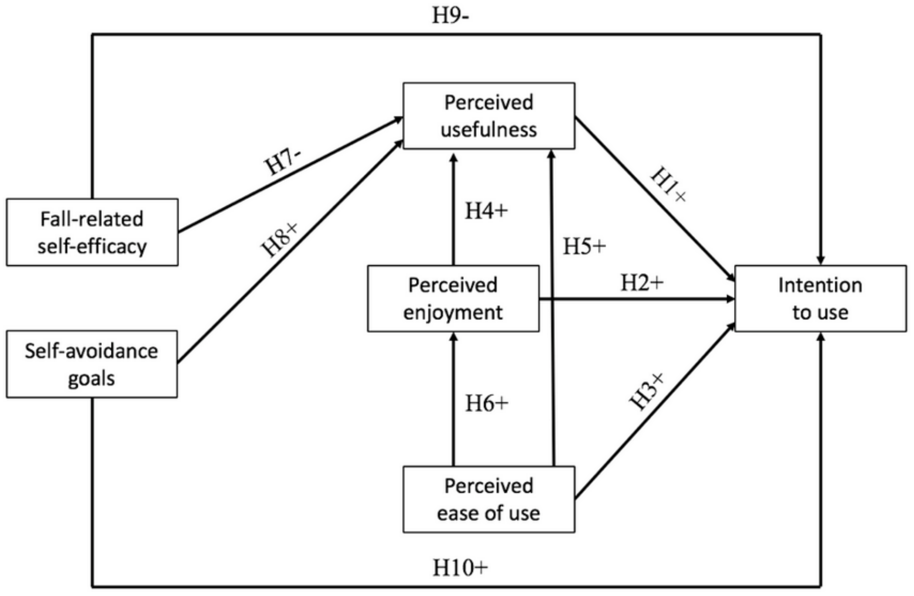
Improving gait and postural control through physical activity training interventions has been shown to be effective for preventing falls [4,5]. Novel applications of technology that promote these interventions appear promising. A number of recent reviews and meta-analyses revealed that new technologies (eg, virtual reality [VR], augmented reality, exergames, and artificial intelligence) have opened the door to a new generation of intervention programs designed to prevent the occurrence of falls [6-8]. Recent studies also demonstrated the effectiveness of training programs implemented via new technologies for improving the control mechanisms involved in balance [9,10] and goal-directed locomotion [11,12] in older adults. For instance, specific exergames have been designed to train balance control, focusing on either the static control of the center of mass or its control while performing precise, rapid, and well-directed steps in balance-threatening situations [13,14]. They allow the cognitive and motor demands of the training conditions to be finely controlled during balance control tasks. However, falls most often occur during everyday walking in complex environments. Therefore, interventions that combine gait training and virtual environments strewn with obstacles have been suggested to be more effective for decreasing fall risk than either classic or balance training interventions delivered via commercial exergames [15].

These programs based on virtual environments provide more realistic/ecological stimulations of physical, cognitive, and sensory resources [16], thereby helping to improve adaptive behavior and prevent falls during daily living tasks [11,12,17]. Even if they pursue the same goal (ie, the development of adaptive behaviors), these programs differ from each other in several aspects, including the technological supports (eg, VR helmet versus projection screen), proposed procedure (more or fewer trials distributed over a longer or shorter period), and tasks used to optimize adaptive capacities [18]. It is therefore not surprising that the results obtained in these studies are sometimes different or even contradictory, and it's often not possible to determine the precise cause of the conflicting results. One explanation for these discrepancies could lie in the different levels of acceptance of the prevention programs. Despite their intrinsic value, as previously identified, technology-based intervention programs must first be accepted by older adults because if they do not see the need for a technology or recognize its usefulness, they are highly unlikely to start using it [19] and may in turn refuse to commit to the training program. Moreover, they may still participate in the program but with less confidence and interest and more psychological discomfort because of their difficulty in accepting the technology to be used. Thus, in view of the potential effectiveness of VR as a rehabilitation tool for fall prevention, it is important to ensure that this tool is well accepted by the target population of the training interventions. This idea has been confirmed and emphasized by a number of recent meta-analyses and reviews [8,20]. Accordingly, the aim of this study was to assess the acceptance of a VR headset (VRH; eg, HTC Vive [HTC Corporation], Oculus Rift [Facebook Technologies, LLC]) by older adults who will use it during a fall-prevention training program. The study of acceptance is thus considered a prerequisite, and will make it possible to refine the analysis of the effectiveness of training interventions, in particular by comparing respondents and nonrespondents.

Acceptance of Technology and the VRH

Acceptance of a particular technology may be defined as the psychological determinants of the behavioral intention to use the technology without ever having experienced it and/or after its actual use [21,22]. Acceptance is based on two well-known models in the social psychology literature, namely the theory of reasoned action and the theory of planned behavior, both of which postulate that behavioral intention and effective behavior are determined by attitudes and representations [23,24]. In other words, both the intention to participate and the effective participation in a technology-based training program may be determined by attitudes and representations toward the training program itself, but also toward the technology used in the training program. If attitudes and representations toward the technology are negative, there is a high probability that individuals will not participate in the training program based on this technology despite its potential and validated benefits. In line with these theories, the technology acceptance model (TAM) [25-27] is the most frequently used theoretical framework to study acceptance of technologies in general and of eHealth technologies in particular [28]. The TAM has highlighted that perceived usefulness of a technology and its perceived ease of use are positive predictors of behavioral intention to use the technology, which is itself a predictor of its actual use [25-27]. Perceived enjoyment may also contribute to the acceptance of technology [29]. This last factor is particularly relevant when examining acceptance of VR technologies, which are often considered hedonic [30], especially when they are being used to prevent fall occurrence among older adults [31]. The four components of the TAM (perceived usefulness, perceived ease of use, perceived enjoyment, and intention to use) and their relationships are represented in Figure 1.

Figure 1. The hypothesized model in the study.



Control variables: gender, age, level of education, financial status

However, studies using the TAM to examine acceptance of immersive VR are not very frequent [32], and those that focus on acceptance of VRHs by older adults are even more rare. For instance, acceptance of VRHs by elderly people was considered neutral before use (neither positive nor negative) but became positive after a first use, as compared with a control group of elderly people exposed to time-lapse videos presented on a computer [33]. Older adults showed high acceptance of VRHs after using the devices for 6 weeks, while perceived usefulness, perceived ease of use, and perceived enjoyment had positive effects on the intention to use VRHs [34]. The same pattern of results was found with qualitative investigations and focus groups [35]. However, to the best of our knowledge, acceptance before a first use of a VRH designed to reduce older adults' risk of falling through their use in intervention programs has not yet been studied. The fact that the positive influence on fall prevention of training programs using VRHs has been validated [10] reinforces the relevance of studying acceptance of such a VRH. Indeed, older adults who have a low acceptance score also have a very high probability of never adopting this kind of VRH despite its objective usefulness to maintain and improve their functional capacities, but this hypothesis has not yet been tested. This study is a step in this direction. The aim of this study was to examine if the TAM variables predicted older adults' intention to use a VRH in an intervention program designed to prevent falls.

Although the TAM is a validated and widely used model, it often needs to be extended and upgraded to increase its explanatory power in eHealth [21,28]. One of the possibilities highlighted by Venkatesh et al [36] was to examine external variables of interest (ie, antecedents) that may influence perceived usefulness, perceived ease of use, and perceived enjoyment. This was the second aim of our study. Fall-related self-efficacy and self-avoidance goals were two promising psychological variables to investigate because they may be related to older adults' investment in exercise intervention programs intended to prevent falls.

Fall-Related Self-Efficacy and Self-Avoidance Goals

Identifying elderly people's fear of falling is worthwhile to optimize prevention because interventions targeted toward older adults with high levels of fear of falling may help decrease the risk that a fall will occur [37]. Moreover, older people often compensate for their fear of falling by being less physically active (ie, reducing frequency and duration of mobility) [38], which could paradoxically limit their involvement in technology-based training programs intended to prevent fall occurrence. Fear of falling may be assessed by fall-related self-efficacy, which is the perceived level of confidence that an individual experiences while performing several activities of daily living without falling, such as taking a bath or getting out of bed [39,40]. The general construct of self-efficacy has already been used with an elderly sample as an antecedent of the main variables of the TAM (eg, gerontechnology self-efficacy [41]), but to date, fall-related self-efficacy has not been used as an external variable with the potential to influence the main variables of the TAM. However, this variable seems to be relevant here because this study focuses on older adults for whom fall-related self-efficacy tends to decrease with age [39] and because we examine the acceptance of a VRH whose purpose is precisely to reduce the risk of falling through its use in intervention programs. We expected fall-related self-efficacy to be a negative predictor of perceived usefulness and intention to use this VRH. This expected result would highlight that elderly people with lower fall-related self-efficacy are more likely to find this VRH useful and to have the intention to use it because the purpose of this VRH addresses their concerns about fear of falling.

From this perspective, another psychological variable, namely self-avoidance goals, was of interest. Among several achievement goals that may be pursued in the physical activity domain, self-avoidance goals focus on participating in a physical activity to avoid physical regression [42]. In the sport context, self-avoidance goals are prevalent among older athletes on the downside of their careers [42], but these kinds of goals are also overrepresented in older adults whatever the context because elderly individuals expect and encounter more losses in resources than gains [43]. Consequently, maintaining and preserving their current resources is often the main goal of older adults, especially in life domains such as health and physical activity, in which resource depletion may be easily seen [44]. Adopting self-avoidance goals in the physical domain (ie, striving to avoid physical regression) may have a protective role in fighting the effects of aging because it may encourage older adults to be physically active to maintain their level of physical fitness, which may therefore have many positive consequences on their health [45]. Thus, to avoid physical decline, older adults with self-avoidance goals will certainly have more desire to participate in training programs intended to prevent fall occurrence. While some theories of motivation have already been used with the TAM (eg, self-determination theory [46]), achievement goal theory has not yet been investigated. However, endorsing self-avoidance goals may have a positive influence on the adoption of eHealth technologies because the purpose of these technologies may match with the purpose of older adults' self-avoidance goals (ie, avoiding physical regression). This assumption remains speculative, but we may reasonably believe that older adults' self-avoidance goals in the physical domain could be positive predictors of the perceived usefulness and intention to use a VRH designed to reduce older adults' risk of falling through its use in intervention programs.

Objectives and Hypotheses

To the best of our knowledge, among the few studies that have investigated the acceptance of VRHs in older adults, no study has focused on the use of this technological device for fall prevention [33,34]. Thus, the first aim of this study based on the TAM was to examine older adults' acceptance of a VRH specifically designed to reduce older adults' risk of falling through its use in intervention programs. The second aim of the study was to investigate the potential predictive role of two psychological variables (fall-related self-efficacy and self-avoidance goals) on the perceived usefulness and intention to use such a device. These variables have not yet been used to extend the TAM, but they appear to be relevant in the context of eHealth technologies designed for fall prevention.

Based on the previous literature review focusing on the TAM, we first hypothesized that older adults' intention to use a VRH specifically designed for fall prevention would be positively predicted by perceived usefulness (H1), perceived enjoyment (H2), and perceived ease of use (H3). Moreover, we expected perceived usefulness to be positively predicted by perceived enjoyment (H4) and perceived ease of use (H5). We also expected perceived enjoyment to be positively predicted by perceived ease of use (H6). Based on the literature reviews focusing on fall-related self-efficacy and self-avoidance goals, we hypothesized that perceived usefulness would be negatively predicted by fall-related self-efficacy (H7) and positively predicted by self-avoidance goals (H8). Finally, we expected intention to use to be negatively predicted by fall-related self-efficacy (H9) and positively predicted by self-avoidance goals (H10). All hypotheses and the model tested in the present study are summarized in Figure 1.

Methods

Participants and Procedure

The study sample included 271 French volunteers (171 women, 100 men) aged 65 to 84 years (mean age 73.69 years, SD 6.37 years). Participants were recruited during the last 3 months of 2019 and met the following eligibility criteria: (1) aged 65 years or older, (2) able to walk without a walking aid, and (3) had never used a VRH before. They voluntarily and anonymously filled out the questionnaire containing focal constructs and demographic information that may influence acceptance of VRH (gender, age, level of education, and financial status). Before completing the questionnaire and following the procedure frequently used with older adults [47], participants had to read a short text accompanied by photos. The short text described a VRH (ie, a helmet-mounted display that completely covers the eyes for an immersive 3-dimensional experience) and its specific purpose in the study (ie, to prevent older adults' falls through its use in exercise intervention programs). Five photos were included beside the text showing a VRH alone, a man wearing a VRH, a virtual scene, an elderly lady in a virtual scene wearing a VRH, and an older man in another virtual scene. This procedure was used because many older adults do not necessarily know what a VRH looks like or what it may be used for.

The questionnaire was completed in individual paper-based sessions. Volunteers gave their consent and were given the opportunity to stop their participation at any time during the study without any consequences. The study was conducted in accordance with the Declaration of Helsinki and met the requirements of the Commission Nationale de l'Informatique et des Libertés (n 2004-801). Table 1 summarizes the participants' demographic characteristics.

Table 1. Participants' demographic characteristics (N=271).

Characteristics	Values, n (%)
Gender	
Women	171 (63.1)
Men	100 (36.9)
Age (years)	
65-74	164 (60.5)
75-84	90 (33.2)
≥85	17 (6.3)
Level of education	
No diploma	18 (6.6)
Low-level diploma	71 (26.2)
Medium-level diploma	62 (22.9)
High-level diploma	27 (10.0)
Short graduate diploma	40 (14.8)
Long graduate diploma	53 (19.6)
Financial status	
Adequate financial resources	102 (37.6)
Adequate financial resources, except for unforeseen circumstances	80 (29.5)
Scarce financial resources	53 (19.6)
Lack funding	36 (13.3)

Measures

Perceived usefulness, perceived ease of use, perceived enjoyment, and intention to use were assessed through a total of 12 items focusing on the VRH that was designed to reduce older adults' risk of falling through physical training. These items were adapted from two studies focusing on the acceptance of technology to improve health [48] and on the acceptance of VR hardware [30]. Participants responded to the 3 items used for each variable on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Actual use, which is a component of the TAM [25], was not investigated because actual use was an exclusion criterion in this study. Instead, we wanted to examine the behavioral intention to use a VRH for preventing falls without having experienced the device.

Fall-related self-efficacy was assessed with the 14-item Modified Falls Efficacy Scale [39], validated in French [40]. Participants rated their level of confidence in performing several activities of daily living (eg, getting dressed and undressed, going up or down stairs) without falling on an 11-point scale ranging from 0 (not at all confident) to 10 (completely confident). The mean of all 14 items represents the fall-related self-efficacy score, with a higher score indicating greater confidence.

Self-avoidance goals were assessed with the corresponding subscale of the 3×2 Achievement Goal Questionnaire for Sport [42], adapted to physical activity. Participants answered the 3 items (eg, "When I do physical activity, my goal is to avoid doing worse than I usually do") on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

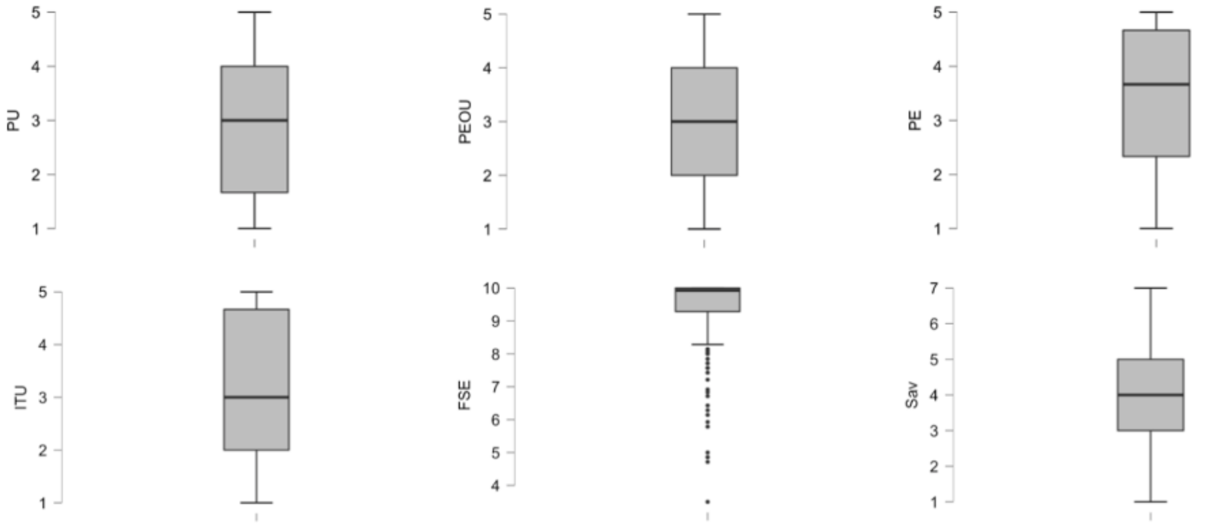
Internal consistency was good for each variable, with McDonald omegas ranging from 0.89 to 0.96. McDonald omegas were used instead of Cronbach alphas because the latter have the tendency to over- or underestimate reliability [49]. McDonald omegas for each variable, descriptive statistics, and the complete list of items are provided in Table 2.

Most of the scales' standard deviations were found to be very high, except for fall-related self-efficacy. Figure 2 shows the boxplot for each variable, highlighting that participants' responses on items assessing perceived usefulness, perceived ease of use, perceived enjoyment, intention to use, and self-avoidance goals were scattered across the entire Likert scale.

Table 2. Descriptive statistics, internal consistency, and items for each variable.

Variables	Mean (SD)	McDonald omega
Perceived usefulness	2.80 (1.25)	0.95
I believe using this VRH would be useful for improving my health.		
I believe using this VRH would enable me to improve my health.		
I believe using this VRH would help me improve my health.		
Perceived ease of use	3.10 (1.27)	0.92
I believe using this VRH would be clear and understandable.		
I would find this VRH easy to use.		
I believe using this VRH would be easy for me.		
Perceived enjoyment	3.42 (1.39)	0.93
I believe I would find using this VRH enjoyable.		
I believe I would have fun using this VRH.		
I believe I would find using this VRH exciting.		
Intention to use	3.15 (1.42)	0.95
Assuming I had access to this VRH, I would like to use it.		
Assuming I had access to this VRH, I would intend to use it.		
Assuming I had access to this VRH, I predict that I would use it.		
Fall-related self-efficacy	9.41 (1.07)	0.96
Getting dressed and undressed/preparing a simple meal/taking a bath or a shower/getting in or out of a chair/getting in or out of bed/answering the door or telephone/walking around the inside of the house/reaching into cabinets or closets/light housekeeping/simple shopping/using public transportation/crossing roads/light gardening or hanging out the washing/going up or down stairs		
Self-avoidance goals	3.57 (1.39)	0.89
When I do physical activity, my goal is to avoid having worse results than I had previously.		
When I do physical activity, my goal is to avoid doing worse than I usually do.		
When I do physical activity, my goal is to avoid being less effective compared with my usual level of performance.		

Figure 2. Boxplots for all variables. FSE: fall-related self-efficacy; ITU: intention to use; PE: perceived enjoyment; PEOU: perceived ease of use; PU: perceived usefulness; Sav: self-avoidance goals.



Data Analysis

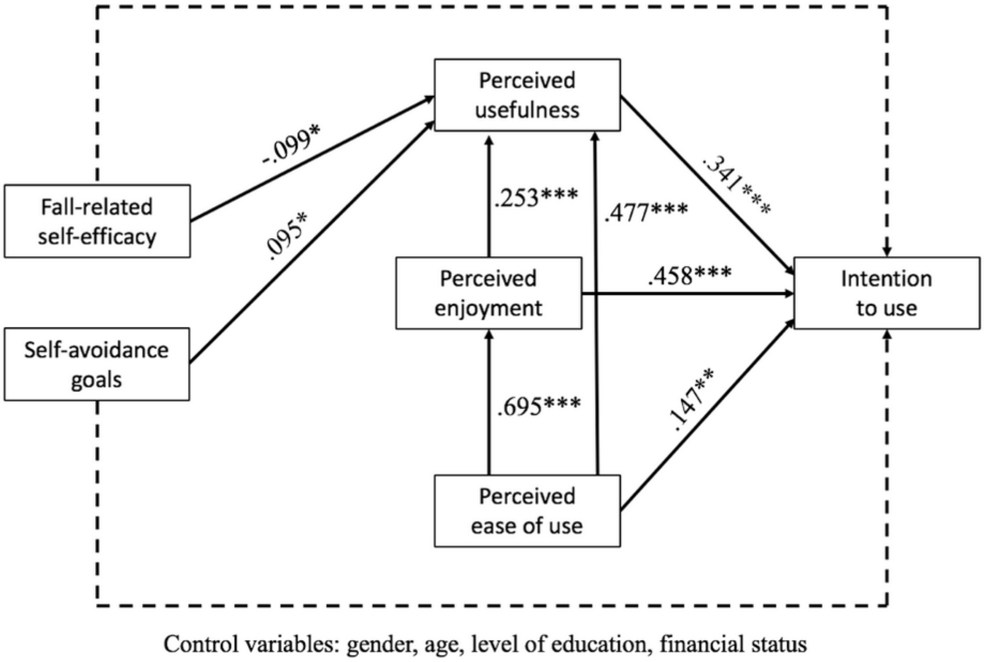
To evaluate the model's fit [50,51], we used the χ^2/df ratio (value ≤3), the comparative fit index (CFI; value ≥0.90), the Tucker-Lewis index (TLI; value ≥0.90), the root mean square error of approximation (RMSEA; value ≤0.08), and the standardized root mean square residual (SRMR; value ≤0.08). A structural equation modeling (SEM) analysis using the maximum likelihood estimations [52] was conducted to test the previous hypotheses (see Figure 1) with the JASP software (version 0.10). Gender, age, level of education, and financial status were entered into the model to control for these variables. According to Kline [53], a typical sample size in studies using SEM is approximately 200 participants. This study met the sample size requirement.

Results

The first results of model fit approached the expected requirements (χ^2[362, N=271]=1124.93, *P*<.001, χ^2/df=3.11, CFI=0.91, TLI=0.90, RMSEA=0.087, and SRMR=0.044), but remained insufficient due to the χ^2/df value above 3 and the value of RMSEA above 0.08. Following the procedure of Kaplan [54] when the initial model did not provide an adequate fit to the data, we examined the modification indices suggested by the statistical software. The modification indices proposed that adding an error covariance between items 10, 11, and 12 of the Modified Falls Efficacy Scale would improve the model's fit. This suggestion was theoretically possible and was used in the present study because these 3 items measured the same latent variable (fall-related self-efficacy). Subsequent analysis revealed that all of the fit statistics met the criteria for an acceptable fitting model, especially the RMSEA and the χ^2/df, whose values were below the threshold of 0.08 and 3, respectively [50,51]: χ^2(359, N=271)=988.95, *P*<.001, χ^2/df=2.75, CFI=0.93, TLI=0.92, RMSEA=0.079, and SRMR =0.042.

The results of the SEM analysis indicated that intention to use the VRH designed to reduce older adults' risk of falling through its use in intervention programs was positively predicted by perceived usefulness (H1 supported; *P*<.001), perceived enjoyment (H2 supported; *P*<.001), and perceived ease of use (H3 supported; *P*=.01), but was not significantly predicted by fall-related self-efficacy (H9 not supported; *P*=.67) or self-avoidance goals (H10 not supported; *P*=.58). Perceived usefulness was positively predicted by perceived enjoyment (H4 supported; *P*<.001), perceived ease of use (H5 supported; *P*<.001), and self-avoidance goals (H8 supported; *P*=.03), and negatively predicted by fall-related self-efficacy (H7 supported; *P*=.03). Finally, perceived enjoyment was positively predicted by perceived ease of use (H6 supported; *P*<.001). All standardized path coefficients are presented in Figure 3.

Figure 3. Validated structural model with standardized path coefficients. Dotted lines indicate nonsignificant paths. **P*<.05, ***P*<.01, ****P*<.001.



Ancillary analyses showed that gender (women=–1, men=1) was a negative predictor of perceived usefulness (*P*=.04) and a positive predictor of perceived ease of use (*P*<.001). Age was found to be a negative predictor of both perceived enjoyment (*P*=.002) and perceived ease of use (*P*<.001). Finally, level of education was a positive predictor of perceived enjoyment (*P*=.04). These results are not represented in Figure 3 because they were not the most relevant ones relative to the main purposes of the study.

Discussion

Main Findings

Based on the TAM, the main purpose of this study was to examine acceptance among older adults of a VRH designed to prevent falls before its first use. The results provided interesting information. First, they highlighted that intention to use this device was positively predicted by perceived usefulness, perceived ease of use, and perceived enjoyment, which had been self-reported by older adults before using this VRH. These findings validate the TAM—for the first time to our knowledge—in the context of the intention to use a VRH to prevent falls through its use in intervention programs. Second, results also showed that fall-related self-efficacy and self-avoidance goals were negative and positive predictors, respectively, of the perceived usefulness of this VRH, providing external variables that had not yet been discussed in the TAM's literature investigating this specific device.

The results of the SEM analysis showed that the most powerful predictor of intention to use the VRH was perceived enjoyment. High perceived enjoyment resulted in high intention to use the VRH. While perceived usefulness was the strongest predictor in most of the previous TAM studies [25,29], recent studies have highlighted that technologies based on virtual environments are considered hedonic technologies [55], and perceived enjoyment therefore becomes a strong predictor—possibly the strongest one [30]—of intention to use them. Perceived enjoyment was also recently found to be a relevant variable to take into account when studying older adults' acceptance of VRHs [32,34]. Perceived usefulness also strongly positively predicted intention to use the VRH for preventing falls. This result was not surprising because (1) it was in line with the theoretical foundations of the TAM postulating that users tend to adopt a technology when they consider it useful [25,29], (2) it corresponds to the results commonly identified in the literature with various technologies, (3) perceived usefulness is a major variable for acceptance of utilitarian technologies [56], such as the VRH used for fall prevention in this study, (4) Peek et al [57] showed that the perceived usefulness of a technology identified by older adults is crucial for preimplementation acceptance (ie, when a technology has not yet been tried), and (5) perceived usefulness was already found to positively predict intention to use VRHs among adults [30] and older adults [32,34]. Finally, intention to use the VRH designed to prevent falls through its use in intervention programs was positively predicted by its perceived ease of use. When the VRH was perceived to be easy to use, older adults were more intent on using it. This result is consistent with the TAM, which postulates that users' adoption of a technology depends on the level of difficulty associated with its use [25,29]. When the VRH used for fall prevention is considered simple and easy to understand by older adults, their intention to use it increases. While perceived ease of use was not the strongest predictor of intention to use this VRH in this study, it still remains a significant predictor that seems to be particularly relevant with older adults who are not necessarily familiar with such technologies [47]. In sum, this study validated, for the first time to our knowledge, the suitability of the TAM for investigating the acceptance of a VRH designed to prevent falls. Acceptance of the VRH is one of the conditions for the success of VRH-based training programs intended to prevent fall occurrence because if an older adult does not accept the device before using it, their likelihood of participating in the training program may decrease quite significantly [58]. Moreover, acceptance during the postimplementation stage (when users have already experienced a technology) may also be a condition of the success of technology-based training programs by increasing the motivational process necessary to maintain and improve participation throughout the program. However, studying acceptance after use was not a purpose of this study.

A second set of results showed that fall-related self-efficacy (ie, the perceived level of confidence of an individual when performing daily activities without falling) was a negative predictor of perceived usefulness of the VRH designed for fall prevention (ie, less confidence leads to more perceived usefulness). This result is not surprising because older adults who have a low fall-related self-efficacy develop a high fear of falling and reduce their activities [40]. Consequently, an intervention program using a VRH designed to prevent falls may seem more useful for older adults with a low fall-related self-efficacy than for those with a high self-efficacy, which explains the negative prediction found in this study. This is of particular interest in the context of health prevention because identifying fall-related self-efficacy may make it possible to propose such VRH-based interventions to older adults who probably need it most. Agreeing to participate in VR training programs would be a first step in helping to reduce the occurrence of falls and their negative consequences among older adults [37]. The results also showed that self-avoidance goals (ie, participating in a physical activity to avoid physical regression) were positive predictors of perceived usefulness of the VRH among older adults. The VRH designed to prevent falls through intervention programs based on physical exercise was found to be more useful to older adults who wanted to avoid physical regression because most of them aimed to maintain and preserve their current resources, especially in the physical domain [44]. The VRH can help them to do so in concrete terms. Finally, we also expected fall-related self-efficacy and self-avoidance goals to be direct predictors of intention to use the VRH. These hypotheses were not supported. Venkatesh [59] identified that in general external variables influence beliefs (perceived usefulness, perceived ease of use, and perceived enjoyment) rather than intention to use, which has been confirmed in this study. In sum, adding external variables (eg, fall-related self-efficacy and self-avoidance goals) as antecedents of perceived usefulness extends the TAM's predictive power because these variables are selected to fit the technology (VRH in this study), context (fall prevention through intervention programs in this study), and users (older adults in this study) specifically investigated in studies about acceptance [32]. However, many other variables should also be examined as potential predictors of technology use by older adults, such as biophysical (eg, cognitive decline), psychological (eg, willingness to remain independent), and contextual (eg, financial means) factors [41,58].

Current Limitations and Directions for Future Studies

This study had some limitations that might be addressed in future research. First, the participants' responses were self-reported and may have been subject to social desirability [60]. Although the completion of the questionnaires was anonymous, older adults may not want to reveal that they are uncomfortable with new technologies. However, this risk remained limited in this study because the average scores for each variable of the TAM were in the middle of the scales and were not very high. Second, the external variables we chose for this study (fall-related self-efficacy and self-avoidance goals) were psychological rather than contextual variables. History of falls (ie, any fall event experienced by older adults during a specified period of time) may be a relevant variable to include in future studies because history of falls was shown to influence older adults' behaviors in VR training [61]. Investigating other variables from the unified theory of acceptance and use of technology [26,36] might also be relevant to a better understanding of acceptance of the VRH used to prevent falls. For instance, the social influence of older adults' companions and facilitating conditions (eg, having the necessary knowledge to use technologies and being able to get help) may be relevant variables to investigate, all the more so since a questionnaire in French has very recently been validated in young adults [21] and may be relatively easily adapted to older adults. Third, the psychological antecedents of the intention to use the VRH in this study were investigated without having experienced VRHs. Assessing older adults' acceptance of this device after a first use would be interesting. Huygelier et al [33] have shown that older adults' acceptance of VRHs increased after a first use lasting a few minutes compared with a control group. Accordingly, it can be hypothesized that the pattern would be similar for the VRH specifically designed for fall prevention. Finally, in future research conducted in our laboratory and elsewhere, the VRH will be used in long-term training programs (6 weeks at a rate of 3 to 4 sessions/week) to develop adaptive behavior of older adults in perfectly controlled, playful, and evolving contexts. A longitudinal study focusing on the evolution of acceptance throughout the intervention programs might provide relevant information about the dynamics of the acceptance process. Furthermore, long-term technology use by older adults may be influenced by disruptive forces that could cause changes in technology use, which might be investigated through longer longitudinal studies [57].

Conclusions

This study focused on acceptance of a VRH designed for fall prevention through its use in training intervention programs as a possible moderator of the effectiveness of these devices. This study is a first step in this direction. Results showed that intention to use the VRH was predicted by its perceived usefulness, perceived enjoyment, and perceived ease of use. The results also suggested that perceived usefulness of the VRH was negatively predicted by older adults' fall-related self-efficacy (ie, elderly individuals with less confidence found the VRH more useful) and positively predicted by their self-avoidance goals (ie, elderly persons who strived to avoid physical regression found the VRH more useful). Finally, our study allows better understanding of the factors that can influence older adults' acceptance of a VRH designed to prevent falls.

Conflicts of Interest

None declared.

References

1. Berchicci M, Lucci G, Di Russo F. Benefits of physical exercise on the aging brain: the role of the prefrontal cortex. J Gerontol A Biol Sci Med Sci 2013 Nov;68(11):1337-1341 [FREE Full text] [doi: 10.1093/gerona/glt094] [Medline: 23833198]
2. Rubenstein LZ. Falls in older people: epidemiology, risk factors and strategies for prevention. Age Ageing 2006 Sep;35 Suppl 2:ii37-ii41 [FREE Full text] [doi: 10.1093/ageing/afl084] [Medline: 16926202]
3. Gschwind YJ, Eichberg S, Marston HR, Ejupi A, Rosario HD, Kroll M, et al. ICT-based system to predict and prevent falls (iStoppFalls): study protocol for an international multicenter randomized controlled trial. BMC Geriatr 2014 Aug 20;14:91 [FREE Full text] [doi: 10.1186/1471-2318-14-91] [Medline: 25141850]
4. Lord S, Sherrington C, Menz H. Falls in older people: risk factors and strategies for prevention. Cambridge: Cambridge University Press; 2001.
5. Sherrington C, Whitney JC, Lord SR, Herbert RD, Cumming RG, Close JCT. Effective exercise for the prevention of falls: a systematic review and meta-analysis. J Am Geriatr Soc 2008 Dec;56(12):2234-2243. [doi: 10.1111/j.1532-5415.2008.02014.x] [Medline: 19093923]
6. Corregidor-Sánchez AI, Segura-Fragoso A, Rodríguez-Hernández M, Criado-Alvarez JJ, González-Gonzalez J, Polonio-López B. Can exergames contribute to improving walking capacity in older adults? A systematic review and meta-analysis. Maturitas 2020 Feb;132:40-48. [doi: 10.1016/j.maturitas.2019.12.006] [Medline: 31883662]
7. Hamm J, Money AG, Atwal A, Paraskevopoulos I. Fall prevention intervention technologies: A conceptual framework and survey of the state of the art. J Biomed Inform 2016 Feb;59:319-345 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.013] [Medline: 26773345]
8. Neri SG, Cardoso JR, Cruz L, Lima RM, de Oliveira RJ, Iversen MD, et al. Do virtual reality games improve mobility skills and balance measurements in community-dwelling older adults? Systematic review and meta-analysis. Clin Rehabil 2017 Oct;31(10):1292-1304. [doi: 10.1177/0269215517694677] [Medline: 28933612]
9. Fu AS, Gao KL, Tung AK, Tsang WW, Kwan MM. Effectiveness of Exergaming Training in Reducing Risk and Incidence of Falls in Frail Older Adults With a History of Falls. Arch Phys Med Rehabil 2015 Dec;96(12):2096-2102. [doi: 10.1016/j.apmr.2015.08.427] [Medline: 26360975]

10. Phu S, Vogrin S, Al Saedi A, Duque G. Balance training using virtual reality improves balance and physical performance in older adults at high risk of falls. Clin Interv Aging 2019;14:1567-1577 [FREE Full text] [doi: 10.2147/CIA.S220890] [Medline: 31695345]
11. Mirelman A, Rochester L, Maidan I, Del Din S, Alcock L, Nieuwhof F, et al. Addition of a non-immersive virtual reality component to treadmill training to reduce fall risk in older adults (V-TIME): a randomised controlled trial. Lancet 2016 Sep 17;388(10050):1170-1182. [doi: 10.1016/S0140-6736(16)31325-3] [Medline: 27524393]
12. van Ooijen MW, Roerdink M, Trekop M, Janssen TWJ, Beek PJ. The efficacy of treadmill training with and without projected visual context for improving walking ability and reducing fall incidence and fear of falling in older adults with fall-related hip fracture: a randomized controlled trial. BMC Geriatr 2016 Dec 28;16(1):215 [FREE Full text] [doi: 10.1186/s12877-016-0388-x] [Medline: 28031021]
13. Giannouli E, Morat T, Zijlstra W. A Novel Square-Stepping Exercise Program for Older Adults (StepIt): Rationale and Implications for Falls Prevention. Front Med (Lausanne) 2019;6:318 [FREE Full text] [doi: 10.3389/fmed.2019.00318] [Medline: 31993435]
14. Shigematsu R, Okura T, Nakagaichi M, Tanaka K, Sakai T, Kitazumi S, et al. Square-stepping exercise and fall risk factors in older adults: a single-blind, randomized controlled trial. J Gerontol A Biol Sci Med Sci 2008 Jan;63(1):76-82. [doi: 10.1093/gerona/63.1.76] [Medline: 18245764]
15. Mirelman A, Rochester L, Reelick M, Nieuwhof F, Pelosin E, Abbruzzese G, et al. V-TIME: a treadmill training program augmented by virtual reality to decrease fall risk in older adults: study design of a randomized controlled trial. BMC Neurol 2013 Feb 06;13:15 [FREE Full text] [doi: 10.1186/1471-2377-13-15] [Medline: 23388087]
16. Howard MC. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Computers in Human Behavior 2017 May;70:317-327. [doi: 10.1016/j.chb.2017.01.013]
17. de Bruin ED, Schoene D, Pichierri G, Smith ST. Use of virtual reality technique for the training of motor control in the elderly. Some theoretical considerations. Z Gerontol Geriatr 2010 Aug;43(4):229-234. [doi: 10.1007/s00391-010-0124-7] [Medline: 20814798]
18. Corregidor-Sánchez AI, Segura-Fragoso A, Criado-Álvarez JJ, Rodríguez-Hernández M, Mohedano-Moriano A, Polonio-López B. Effectiveness of Virtual Reality Systems to Improve the Activities of Daily Life in Older People. Int J Environ Res Public Health 2020 Aug 28;17(17):6283 [FREE Full text] [doi: 10.3390/ijerph17176283] [Medline: 32872313]
19. Peek STM, Wouters EJM, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJM. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
20. Skjæret N, Nawaz A, Morat T, Schoene D, Helbostad JL, Vereijken B. Exercise and rehabilitation delivered through exergames in older adults: An integrative review of technologies, safety and efficacy. Int J Med Inform 2016 Jan;85(1):1-16. [doi: 10.1016/j.ijmedinf.2015.10.008] [Medline: 26559887]
21. Hayotte M, Thérouanne P, Gray L, Corrion K, d'Arripe-Longueville F. The French eHealth Acceptability Scale Using the Unified Theory of Acceptance and Use of Technology 2 Model: Instrument Validation Study. J Med Internet Res 2020 Apr 15;22(4):e16520 [FREE Full text] [doi: 10.2196/16520] [Medline: 32293569]
22. Alexandre B, Reynaud E, Osiurak F, Navarro J. Acceptance and acceptability criteria: a literature review. Cogn Tech Work 2018 Jan 24;20(2):165-177. [doi: 10.1007/s10111-018-0459-1]
23. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Englewood Cliffs, NJ: Prentice-Hall; 1980.
24. Fishbein M, Ajzen I. Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley; 1975.
25. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
26. Venkatesh V, Bala H. Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
27. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
28. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A Systematic Review of the Technology Acceptance Model in Health Informatics. Appl Clin Inform 2018 Dec;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
29. Davis FD, Bagozzi RP, Warshaw PR. Extrinsic and Intrinsic Motivation to Use Computers in the Workplace1. J Appl Social Psychol 1992 Jul;22(14):1111-1132. [doi: 10.1111/j.1559-1816.1992.tb00945.x]
30. Manis KT, Choi D. The virtual reality hardware acceptance model (VR-HAM): Extending and individuating the technology acceptance model (TAM) for virtual reality hardware. Journal of Business Research 2019 Jul;100:503-513. [doi: 10.1016/j.jbusres.2018.10.021]
31. Mirelman A, Maidan I, Shiratzky S, Hausdorff J. Virtual Reality Training as an Intervention to Reduce Falls. In: Montero-Odasso M, Camicioli R, editors. Falls and Cognition in Older Persons: Fundamentals, Assessment and Therapeutic Options. Cham: Springer Nature Switzerland AG; 2020:309-321.
32. Sagnier C, Loup-Escande E, Lourdeaux D, Thouvenin I. User acceptance of virtual reality: an extended technology acceptance model. Int J Hum Comput Interact 2020;36(11):993-1007. [doi: 10.1080/10447318.2019.1708612]

33. Huygelier H, Schraepen B, van Ee R, Vanden Abeele V, Gillebert CR. Acceptance of immersive head-mounted virtual reality in older adults. Sci Rep 2019 Mar 14;9(1):4519 [FREE Full text] [doi: 10.1038/s41598-019-41200-6] [Medline: 30872760]
34. Syed-Abdul S, Malwade S, Nursetyo AA, Sood M, Bhatia M, Barsasella D, et al. Virtual reality among the elderly: a usefulness and acceptance study from Taiwan. BMC Geriatr 2019 Aug 19;19(1):223 [FREE Full text] [doi: 10.1186/s12877-019-1218-8] [Medline: 31426766]
35. Roberts AR, De Schutter B, Franks K, Radina ME. Older Adults' Experiences with Audiovisual Virtual Reality: Perceived Usefulness and Other Factors Influencing Technology Acceptance. Clin Gerontol 2019;42(1):27-33. [doi: 10.1080/07317115.2018.1442380] [Medline: 29505343]
36. Venkatesh, Thong, Xu. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
37. Gettens S, Fulbrook P. Fear of falling: association between the Modified Falls Efficacy Scale, in-hospital falls and hospital length of stay. J Eval Clin Pract 2015 Feb;21(1):43-50. [doi: 10.1111/jep.12226] [Medline: 25040834]
38. Wijlhuizen GJ, de Jong R, Hopman-Rock M. Older persons afraid of falling reduce physical activity to prevent outdoor falls. Prev Med 2007 Mar;44(3):260-264. [doi: 10.1016/j.ypmed.2006.11.003] [Medline: 17184828]
39. Hill KD, Schwarz JA, Kalogeropoulos AJ, Gibson SJ. Fear of falling revisited. Arch Phys Med Rehabil 1996 Oct;77(10):1025-1029. [doi: 10.1016/s0003-9993(96)90063-5] [Medline: 8857881]
40. Perrot A, Castanier C, Maillot P, Zitari H. French validation of the modified-falls efficacy scale (M-FES Fr). Arch Gerontol Geriatr 2018;78:233-239. [doi: 10.1016/j.archger.2018.07.001] [Medline: 30025268]
41. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
42. Mascret N, Elliot AJ, Cury F. Extending the 3 × 2 achievement goal model to the sport domain: The 3 × 2 Achievement Goal Questionnaire for Sport. Psychology of Sport and Exercise 2015 Mar;17:7-14. [doi: 10.1016/j.psychsport.2014.11.001]
43. Mustafić M, Freund AM. Age-related differences in evaluating developmental stability. International Journal of Behavioral Development 2013 Jul 10;37(4):376-386. [doi: 10.1177/0165025413490866]
44. Senko C, Freund AM. Are mastery-avoidance achievement goals always detrimental? An adult development perspective. Motiv Emot 2015 Feb 10;39(4):477-488. [doi: 10.1007/s11031-015-9474-1]
45. Cunningham C, O' Sullivan R, Caserotti P, Tully MA. Consequences of physical inactivity in older adults: A systematic review of reviews and meta-analyses. Scand J Med Sci Sports 2020 May;30(5):816-827. [doi: 10.1111/sms.13616] [Medline: 32020713]
46. Dupuy L, Consel C, Sauzéon H. Self determination-based design to achieve acceptance of assisted living technologies for older adults. Computers in Human Behavior 2016 Dec;65:508-521. [doi: 10.1016/j.chb.2016.07.042]
47. Li J, Ma Q, Chan AH, Man SS. Health monitoring through wearable technologies for older adults: Smart wearables acceptance model. Appl Ergon 2019 Feb;75:162-169. [doi: 10.1016/j.apergo.2018.10.006] [Medline: 30509522]
48. Chen TL, Bhattacharjee T, Beer JM, Ting LH, Hackney ME, Rogers WA, et al. Older adults' acceptance of a robot for partner dance-based exercise. PLoS One 2017;12(10):e0182736 [FREE Full text] [doi: 10.1371/journal.pone.0182736] [Medline: 29045408]
49. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 2014 Aug;105(3):399-412. [doi: 10.1111/bjop.12046] [Medline: 24844115]
50. Byrne B. Structural equation modelling with AMOS: Basic concepts, applications, and programming. New York, NY: Routledge; 2010.
51. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
52. Everitt B, Hothorn T. An introduction to applied multivariate analysis with R. Applied spatial data analysis with R. New York, NY: Springer; 2011.
53. Kline R. Principles and practice of structural equation modeling. New York: Guilford Press; 2011.
54. Kaplan D. A Study of the Sampling Variability and z-Values of Parameter Estimates From Misspecified Structural Equation Models. Multivariate Behav Res 1989 Jan 01;24(1):41-57. [doi: 10.1207/s15327906mbr2401_3] [Medline: 26794295]
55. Sun H, Cheng W. The input-interface of Webcam applied in 3D virtual reality systems. Computers & Education 2009 Dec;53(4):1231-1240. [doi: 10.1016/j.compedu.2009.06.006]
56. van der Heijden. User Acceptance of Hedonic Information Systems. MIS Quarterly 2004;28(4):695. [doi: 10.2307/25148660]
57. Peek STM, Luijkx KG, Vrijhoef HJM, Nieboer ME, Aarts S, van der Voort CS, et al. Understanding changes and stability in the long-term use of technologies by seniors who are aging in place: a dynamical framework. BMC Geriatr 2019 Aug 28;19(1):236 [FREE Full text] [doi: 10.1186/s12877-019-1241-9] [Medline: 31462214]
58. Peek STM, Luijkx KG, Rijnaard MD, Nieboer ME, van der Voort CS, Aarts S, et al. Older Adults' Reasons for Using Technology while Aging in Place. Gerontology 2016;62(2):226-237 [FREE Full text] [doi: 10.1159/000430949] [Medline: 26044243]
59. Venkatesh V. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. Information Systems Research 2000 Dec;11(4):342-365. [doi: 10.1287/isre.11.4.342.11872]

60. Phillips DL, Clancy KJ. Some Effects of "Social Desirability" in Survey Studies. American Journal of Sociology 1972 Mar;77(5):921-940. [doi: 10.1086/225231]
61. Dermody G, Whitehead L, Wilson G, Glass C. The Role of Virtual Reality in Improving Health Outcomes for Community-Dwelling Older Adults: Systematic Review. J Med Internet Res 2020 Jun 01;22(6):e17331. [doi: 10.2196/17331] [Medline: 32478662]

Abbreviations

CFI: comparative fit index
RMSEA: root mean square error of approximation
SEM: structural equation modeling
SRMR: standardized root mean square residual
TAM: technology acceptance model
TLI: Tucker-Lewis index
VR: virtual reality
VRH: virtual reality headset

Edited by G Eysenbach; submitted 26.05.20; peer-reviewed by T Cleworth, L van Velsen; comments to author 24.07.20; revised version received 17.09.20; accepted 28.10.20; published 14.12.20.

Please cite as:
Mascret N, Delbes L, Voron A, Temprado JJ, Montagne G
Acceptance of a Virtual Reality Headset Designed for Fall Prevention in Older Adults: Questionnaire Study
J Med Internet Res 2020;22(12):e20691
URL: http://www.jmir.org/2020/12/e20691/
doi:10.2196/20691
PMID:33315019

Original Paper

Physiotherapists' Use of Web-Based Information Resources to Fulfill Their Information Needs During a Theoretical Examination: Randomized Crossover Trial

Cailbhe Doherty[1,2], BSc, PhD; Arash Joorabchi[2,3], BSc, PhD; Peter Megyesi[2], BSc; Aileen Flynn[4], BSc, MSc; Brian Caulfield[2], BSc, MSc, PhD

[1]School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland
[2]Insight Centre for Data Analytics, O'Brien Centre for Science, University College Dublin, Dublin, Ireland
[3]Department of Electronic & Computer Engineering, University of Limerick, Limerick, Ireland
[4]Beacon Hospital, Dublin, Ireland

Corresponding Author:
Cailbhe Doherty, BSc, PhD
School of Public Health, Physiotherapy and Sports Science
University College Dublin
A308, Health Science Building
Belfield
Dublin, D4
Ireland
Phone: 353 17166511
Email: cailbhe.doherty@ucd.ie

Abstract

Background: The widespread availability of internet-connected smart devices in the health care setting has the potential to improve the delivery of research evidence to the care pathway and fulfill health care professionals' information needs.

Objective: This study aims to evaluate the frequency with which physiotherapists experience information needs, the capacity of digital information resources to fulfill these needs, and the specific types of resources they use to do so.

Methods: A total of 38 participants (all practicing physiotherapists; 19 females, 19 males) were randomly assigned to complete three 20-question multiple-choice questionnaire (MCQ) examinations under 3 conditions in a randomized crossover study design: assisted by a web browser, assisted by a federated search portal system, and unassisted. MCQ scores, times, and frequencies of information needs were recorded for overall examination-level and individual question-level analyses. Generalized estimating equations were used to assess differences between conditions for the primary outcomes. A log file analysis was conducted to evaluate participants' web search and retrieval behaviors.

Results: Participants experienced an information need in 55.59% (845/1520) MCQs (assisted conditions only) and exhibited a mean improvement of 10% and 16% in overall examination scores for the federated search and web browser conditions, respectively, compared with the unassisted condition ($P<.001$). In the web browser condition, Google was the most popular resource and the only search engine used, accounting for 1273 (64%) of hits, followed by PubMed (195 hits; 10% of total). In the federated search condition, Wikipedia and PubMed were the most popular resources with 1518 (46% of total) and 1273 (39% of total) hits, respectively.

Conclusions: In agreement with the findings of previous research studies among medical physicians, the results of this study demonstrate that physiotherapists frequently experience information needs. This study provides new insights into the preferred digital information resources used by physiotherapists to fulfill these needs. Future research should clarify the implications of physiotherapists' apparent high reliance on Google, whether these results reflect the authentic clinical environment, and whether fulfilling clinical information needs alters practice behaviors or improves patient outcomes.

(J Med Internet Res 2020;22(12):e19747) doi:10.2196/19747

KEYWORDS

evidence-based medicine; knowledge discovery; information seeking behavior; information dissemination; information literacy; online systems; point-of-care systems; mobile phone

Introduction

Health Care Professionals' Information Seeking Behaviors: A Brief History

The ubiquity of internet-connected smart devices in the clinical setting [1] and the availability of preappraised point-of-care evidence summaries [2,3] have changed how health care professionals seek, consume, and implement information in the care pathway [4]. For instance, a generation now separates the sample of health care professionals who participated in the seminal work of Covell et al [5] in 1985 and their digitally native descendants. That sample (47 medical physicians) reported having 2 clinical questions for every 3 patient visits and preferred textbooks, journals, and drug information indexes to fulfill their information needs [5]. Importantly, these findings are now irreconcilable with modern clinical archetypes as "computered sources were reported to be used least often" [5].

Since then, the delivery of clinically relevant information to the care pathway has progressed to smart devices capable of running stand-alone software apps. At present, an increasing proportion of health care professionals use these devices to inform their practice [1]. This change has been charted in thousands of studies [6] and syntheses of studies [7,8], which have demonstrated the use of web-connected smart devices for fulfilling clinical information needs [5,8-11]

Empirical Research of Health Care Professionals' Information Seeking Behaviors

Unfortunately, the current understanding of information-seeking and utilization behaviors is shaped by suboptimal empirical models. Self-report questionnaires [5,9], interviews [12], and think-aloud [13-15] methodologies may offer a feasible way to estimate how health care professionals perceive their own information-seeking behaviors, their frequency, the topics to which they relate, the resources they use to fulfill these needs, and the strategies they employ to harvest information from these resources. However, these behaviors can now be more directly observed using web server log file analysis [16-18].

To date, researchers have leveraged web logs to evaluate usage patterns for specific websites such as Wikipedia [19,20], PubMed [21,22], UpToDate [23], and web-based institutional health science libraries [16,17]. Typically reported usage metadata include the number of time- or user-dependent hits or sessions on the website [24,25], the most frequently searched topics [18,22], and the number of click-throughs from one section of the website to another [18].

Importantly, these analyses tend to be constrained to individual websites [22] or institutional portals [23], which redirect to a limited group of external websites. These constraints undermine the validity of modern information-seeking paradigms and bely our understanding of the specific digital information resources that are used in an unconstrained web environment to fulfil information needs. Health care professionals in the real world have unbounded web access; this does not reflect the sandboxes of preselected information resources that are evaluated in the available body of research. Furthermore, little work in the web log analysis literature has focused on how these resources can be used to answer clinical questions.

Aims and Objectives of the Study

Therefore, the aim of this study is to evaluate the use of web-based resources for fulfilling clinical information needs. Specifically, our objectives were to conduct a randomized crossover trial, whereby a group of physiotherapists were subjected to a multiple-choice questionnaire (MCQ) examination, which they completed under 3 conditions: (1) assisted via a web browser with unconstrained web access, (2) assisted via a federated search engine "portal" app, and (3) unassisted.

By including both a federated search system and unconstrained web use in a single study design, we sought to evaluate the potentially mediating effects of the access tool on the kinds of resources being used and the time spent doing so. This study aims to address both hypothesis-confirming and hypothesis-generating research questions.

Experimental Hypotheses

The confirmatory hypotheses were as follows:

1. Physiotherapists frequently encounter information needs when presented with simulated clinical questions [5,9-11].
2. Digital information resources available on the web can be used to fulfill these needs, resulting in a higher MCQ examination score in the assisted conditions [26].

The hypotheses that this research may generate relate to the following:

1. The specific web-based resources used by physiotherapists to fulfill their information needs.
2. The rate of answering correctly in the presence or absence of an information need and its relationship with the use or nonuse of a digital information resource.
3. Whether constraining physiotherapists to a limited number of web-based digital information resources (as in the federated search condition) is associated with a difference in the rate with which questions are answered correctly or the time spent in searching for information.

Methods

Ethics

Ethical approval for this randomized crossover trial was granted by the affiliate review board of the institution at which the authors were based (reference: LS-18-25). Study design, conduct, analysis, and results are reported according to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.

Recruitment

Prospective participants included practicing physiotherapists who were recruited with the help of collaborating health care institutions. The study was advertised on the websites of these institutions, and a notice was sent throughout the recruitment period to prospective participants via the "staff bulletin" of the respective institutions. Physiotherapists were chosen for evaluation, as their information behavior has not been well addressed in the research literature to date, despite their prominent role in health care [27].

All individuals met the following inclusion criteria: adults (>18 years of age), currently employed by collaborating health-related institutions and involved with the management and treatment of patients, provided informed consent (which was obtained digitally during authentication with the federated search app).

Protocol

Information needs were induced via the administration of an MCQ examination completed at 3 time points under 3 conditions in random order: (1) assisted by a standard web browser, (2) assisted by a federated search engine, and (3) unassisted.

One investigator (CD) organized an appropriate time and location to conduct all test sessions; to maximize participant retention, testing was conducted at a time and place that suited participants. Short-term rescheduling (ie, within 1 week of the original designated test time) was facilitated in the event of unexpected delays or schedule conflicts.

For the purpose of this study, we defined an information need in the assisted conditions as any instance in which a participant used an assistive technology to access a digital information resource to inform their choice of answer in the MCQ. In the "assisted by a web browser" condition, participants were provided with a web-connected laptop preinstalled with a web browser (Google Chrome). For the "assisted by a federated search engine" condition, participants were required to install and register an account with a free experimental federated search engine called *SciScanner* [28] on their smartphone.

The Multiple-Choice Questionnaire

Each theoretical examination comprised a 20-item MCQ with one best answer, derived from the Physiotherapy Competency Exam (PCE) [29] prepared by the Canadian Alliance of Physiotherapy Regulators and the United States National Physical Therapy Exam (NPTE) [30]. A bank of >1000 questions was compiled for each of the PCE and NPTE examinations [31,32]. During each examination, a different random subset was taken from this question bank for each participant for each of the 3 examinations they completed. The MCQ was administered on a tablet device using a commercially available testing software that administered random subsets of topic-tagged questions. Participants were assigned questions based on their chosen area of practice using these tags. The software also documented the question that was administered, its associated response, whether that response was correct or incorrect, and the time spent answering it [33]. Participants were informed that the examination was negatively marked to discourage guessing; however, they were not informed about the penalty for an incorrect answer. Participants were also informed that time was not to be considered a metric for performance, whereby they would not be penalized for finishing the examination in a longer period. In the "assisted" experimental conditions, participants were informed that they could opt to use their assigned system but did not have to if they did not deem it necessary or if they were sufficiently confident of their answer. In either case, participants were informed that they would not be penalized for using or not using assistance if it was available to them. The examination was proctored by the first author, who manually documented the frequency of information needs (defined by instances where participants chose to use either the standard web browser or the federated search engine for a given examination question).

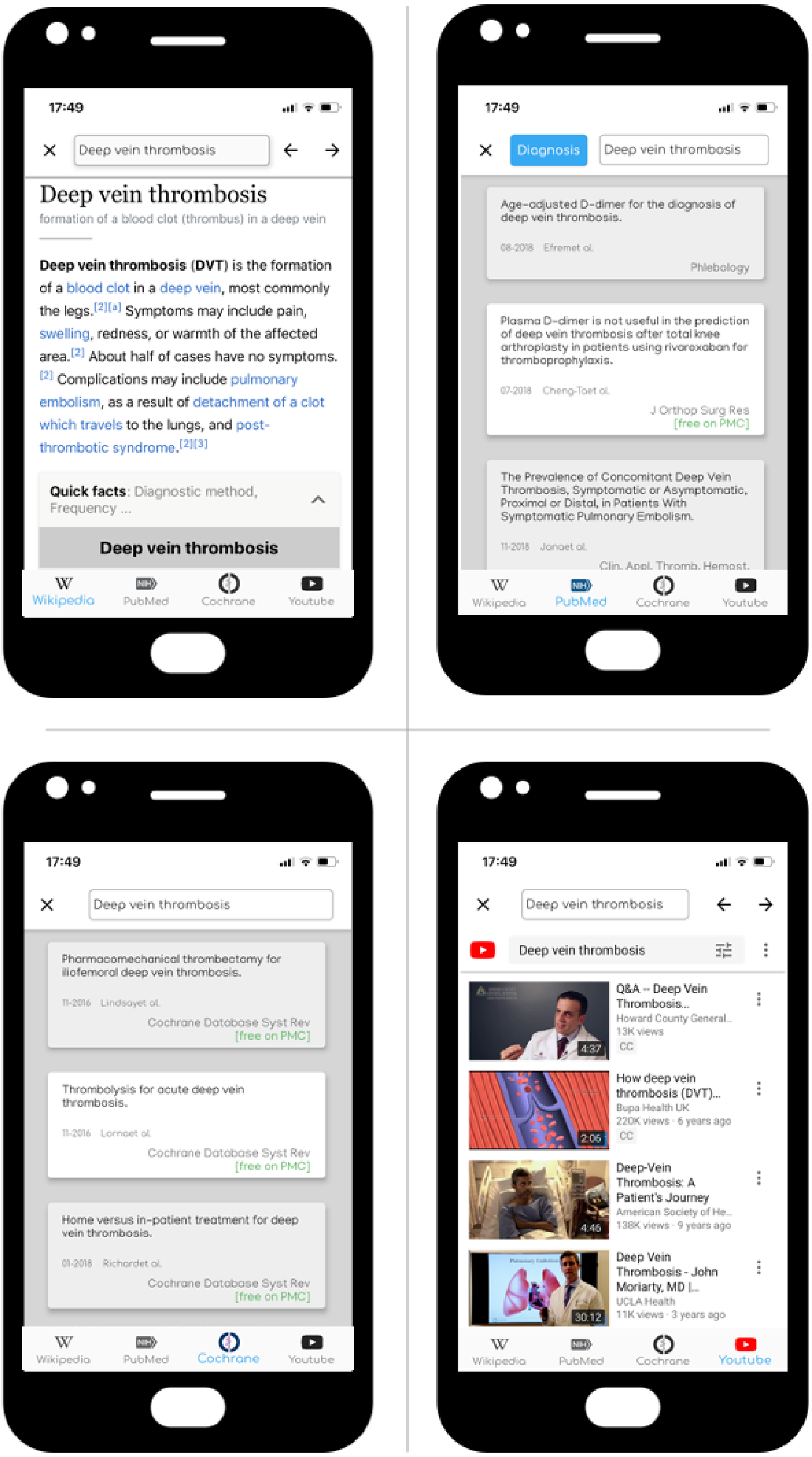
Web and Federated Search Logs

Participants' web logs and federated search logs were captured during the 2 assisted conditions.

Federated Search Logs

Federated searching describes a system that implements a single query concurrently across multiple disparate collections of information. The federated search engine used in this study is a free experimental platform called *SciScanner* that centralizes popular health-information resources [34]. At the time of the study, searches were implemented across a series of authoritative (PubMed and the Cochrane library), nonauthoritative (YouTube), and community - built (Wikipedia) resources in tandem. These were chosen based on prior research identifying the most commonly used, free resources among health care professionals [35-39] and because of the ease with which a simple search string could be used and adapted for each data source while maximizing the consistency of these results [40]. For example, the "Clinical Queries" search strategies were implemented in the system to aid users in finding different types of content on PubMed (eg, related to diagnosis, clinical prediction, or treatment best practices) [41]. The results for a single search query could be accessed for each resource on the app home screen (Figure 1).

Figure 1. Search results for “Deep Vein Thrombosis” on the federated search platform. At the time of the study, the platform queried Wikipedia (top, left), PubMed (“diagnosis” category; top, right), the Cochrane library (bottom, left), and YouTube (bottom, right).



Deep vein thrombosis

formation of a blood clot (thrombus) in a deep vein

Deep vein thrombosis (**DVT**) is the formation of a blood clot in a deep vein, most commonly the legs.[2][a] Symptoms may include pain, swelling, redness, or warmth of the affected area.[2] About half of cases have no symptoms.[2] Complications may include pulmonary embolism, as a result of detachment of a clot which travels to the lungs, and post-thrombotic syndrome.[2][3]

Age-adjusted D-dimer for the diagnosis of deep vein thrombosis.
08-2018 Efremet al.
Phlebology

Plasma D-dimer is not useful in the prediction of deep vein thrombosis after total knee arthroplasty in patients using rivaroxaban for thromboprophylaxis.
07-2018 Cheng-Toet al.
J Orthop Surg Res
[free on PMC]

The Prevalence of Concomitant Deep Vein Thrombosis, Symptomatic or Asymptomatic, Proximal or Distal, in Patients With Symptomatic Pulmonary Embolism.
11-2018 Jonaet al.
Clin. Appl. Thromb. Hemost.

Pharmacomechanical thrombectomy for iliofemoral deep vein thrombosis.
11-2016 Lindsayet al.
Cochrane Database Syst Rev
[free on PMC]

Thrombolysis for acute deep vein thrombosis.
11-2016 Lornoet al.
Cochrane Database Syst Rev
[free on PMC]

Home versus in-patient treatment for deep vein thrombosis.
01-2018 Richardet al.
Cochrane Database Syst Rev
[free on PMC]

Q&A -- Deep Vein Thrombosis... Howard County General... 13K views CC 4:37

How deep vein thrombosis (DVT)... Bupa Health UK 220K views · 6 years ago CC 2:05

Deep-Vein Thrombosis: A Patient's Journey American Society of He... 138K views · 9 years ago 4:46

Deep Vein Thrombosis - John Moriarty, MD |... UCLA Health 11K views · 3 years ago 30:12

Following recruitment and before completion of the protocol, participants were required to download and register an account with the SciScanner federated search platform. This authentication process facilitated the tracking of participants’ account-linked information search and retrieval behaviors. The timestamps of all search events, the accounts associated with these events, the search queries themselves, and the resources used for each search query were logged by the system and could be exported in a comma-separated values (CSV) file format for further statistical analysis.

Web Logs

Web logs were acquired using a web browser extension. This browser extension was installed on the computer provided to participants during the completion of their examination. When opened, the browser loaded a blank homepage; participants were allowed to select their preferred sites and search engines.

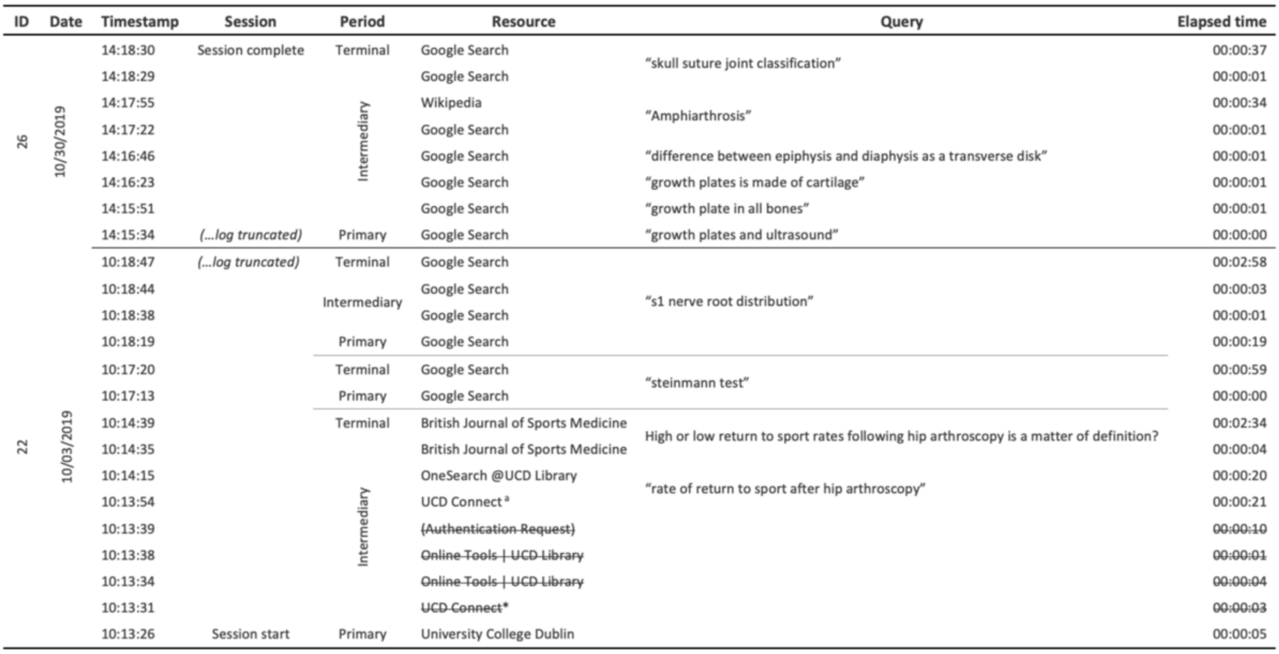
The browser extension captured a similar collection of metrics as those described previously, which were used to quantify information search and retrieval behaviors: search queries, associated timestamps, and the URLs of the webpages that were accessed were logged by the extension. These logs were manually exported in a CSV file format following completion of each examination for further analysis.

Data Preparation

All web and federated search log files were accrued in a single database after participants had completed the test protocol for each "assisted" condition and were filtered to remove setup (eg, log-in events to institutional networks) and duplicate events. All log files were then grouped into sessions based on the timestamps of the first and last log, which were cross-referenced with the recorded date, start time, and duration of the examination. Each session was then manually segmented by the primary author into individual information-seeking events based on search queries and the timestamps of every event; consecutive search queries that were deemed to be thematically related (eg, "sensation to medial tibia," "nerve supply to medial leg," and "saphenous nerve") and those that occurred within a similar timeframe were grouped as a single information-seeking event. Information-seeking events were then partitioned into 3 periods: primary, intermediate, and terminal. The primary period related to the first search that was logged by participants in an information-seeking session or the first search logged in a thematically similar series of consecutive search segments. Terminal periods were defined as the last search logged by participants in an information-seeking session or the last search logged in a thematically similar series of consecutive search segments. Intermediate periods included consecutive logs between the primary and terminal periods.

For each period of an information-seeking event, the resources that were accessed were identified. This was automatically recorded in the federated search logs by design and was determined based on the website URL for the web logs. The number of hits and the accumulated time for each hit were then determined for each resource for the entire test cohort. Segmentation, partitioning, and preparation for data analysis are displayed in Figure 2.

Figure 2. An example collection of search sessions segmented into primary, intermediate, and terminal search periods. Strikethrough text depicts that events would have been removed during data preparation, as they coincided with log-in and authentication to an institutional website.



ID	Date	Timestamp	Session	Period	Resource	Query	Elapsed time
26	10/30/2019	14:18:30	Session complete	Terminal	Google Search	"skull suture joint classification"	00:00:37
		14:18:29		Intermediary	Google Search		00:00:01
		14:17:55			Wikipedia	"Amphiarthrosis"	00:00:34
		14:17:22			Google Search		00:00:01
		14:16:46			Google Search	"difference between epiphysis and diaphysis as a transverse disk"	00:00:01
		14:16:23			Google Search	"growth plates is made of cartilage"	00:00:01
		14:15:51			Google Search	"growth plate in all bones"	00:00:01
		14:15:34	*(…log truncated)*	Primary	Google Search	"growth plates and ultrasound"	00:00:00
22	10/03/2019	10:18:47	*(…log truncated)*	Terminal	Google Search	"s1 nerve root distribution"	00:02:58
		10:18:44		Intermediary	Google Search		00:00:03
		10:18:38			Google Search		00:00:01
		10:18:19		Primary	Google Search		00:00:19
		10:17:20		Terminal	Google Search	"steinmann test"	00:00:59
		10:17:13		Primary	Google Search		00:00:00
		10:14:39		Terminal	British Journal of Sports Medicine	High or low return to sport rates following hip arthroscopy is a matter of definition?	00:02:34
		10:14:35		Intermediary	British Journal of Sports Medicine		00:00:04
		10:14:15			OneSearch @UCD Library	"rate of return to sport after hip arthroscopy"	00:00:20
		10:13:54			UCD Connect^a		00:00:21
		10:13:39			~~(Authentication Request)~~		~~00:00:10~~
		10:13:38			~~Online Tools | UCD Library~~		~~00:00:01~~
		10:13:34			~~Online Tools | UCD Library~~		~~00:00:04~~
		10:13:31			~~UCD Connect*~~		~~00:00:03~~
		10:13:26	Session start	Primary	University College Dublin		00:00:05

Sample Size

Previous research on knowledge acquisition using web-based resources among medical trainees identified a 15% difference between assisted and nonassisted assessment scores (Cohen *d*=0.86) [21,26]. On this basis, a sample of 18 pairs of physiotherapists was deemed sufficient to achieve experimental power. To account for a 5% (2 of 36) potential dropout, 38 participants (19 pairs) were recruited; however, there were no dropouts, so all participants were included in the analysis.

Outcomes

We adopted a 3-tiered analysis paradigm for our outcomes. The first tier was related to the MCQ results (examination-level analysis), the second to the individual questions of the MCQ (question-level analysis), and the third was related to the log files recorded in instances where there was an information need (log-level analysis; assisted conditions only).

Examination-Level Analysis

The salient outcomes for the examination-level analysis included participants' overall MCQ scores and their examination times for each condition. In addition, the total number of information needs was a salient outcome for examinations completed in the "assisted" conditions. Total examination scores were computed without implementing any negative marking.

Question-Level Analysis

The salient outcomes for the question-level analysis included the presence or absence of an information need for each question, the time spent answering each question, and whether the question was answered correctly or incorrectly.

Log-Level Analysis

The salient outcomes for the log-level analysis included the number of hits, the time spent per hit, and the digital information resource that was used.

Statistical Analysis

The demographics of the participating sample were represented using descriptive statistics.

Examination-Level Analysis

Generalized estimating equations (GEEs) were used to assess differences in examination scores (assisted with a web browser, assisted with a federated search engine, and unassisted conditions) and information needs (assisted with a web browser and assisted with federated search engine conditions) with examination time included as a covariate using an exchangeable correlation structure. The model was corrected for dependent observations by including the participants' identifying code as a subject effect. The *a priori p* value for this analysis was set at *P*<.05.

Question-Level Analysis

A separate model GEE was defined to evaluate the effect of different conditions on individual answers in the presence and absence of an information need with question time included as a covariate in the model. The model was corrected for dependent observations by including the participants' identifying code as a subject effect. The *a priori p* value for this analysis was set at *P*<.05.

Log-Level Analysis

Log data for all web-based and federated search–based information resources were represented using means with SDs or medians with IQR, where appropriate, for each period of the information trail. GEEs were used to assess the differences in the time spent per "hit" in each of the 3 periods of the information trail (primary, intermediate, and tertiary) for each of the assisted conditions using an exchangeable correlation structure. This model was corrected for dependent observations by including the participants' identifying code as a subject effect. The *a priori p* value for this analysis was set at *P*<.05.

All data analyses were performed using Statistical Package for the Social Sciences version 26 (IBM Corp) and Microsoft Excel.

Results

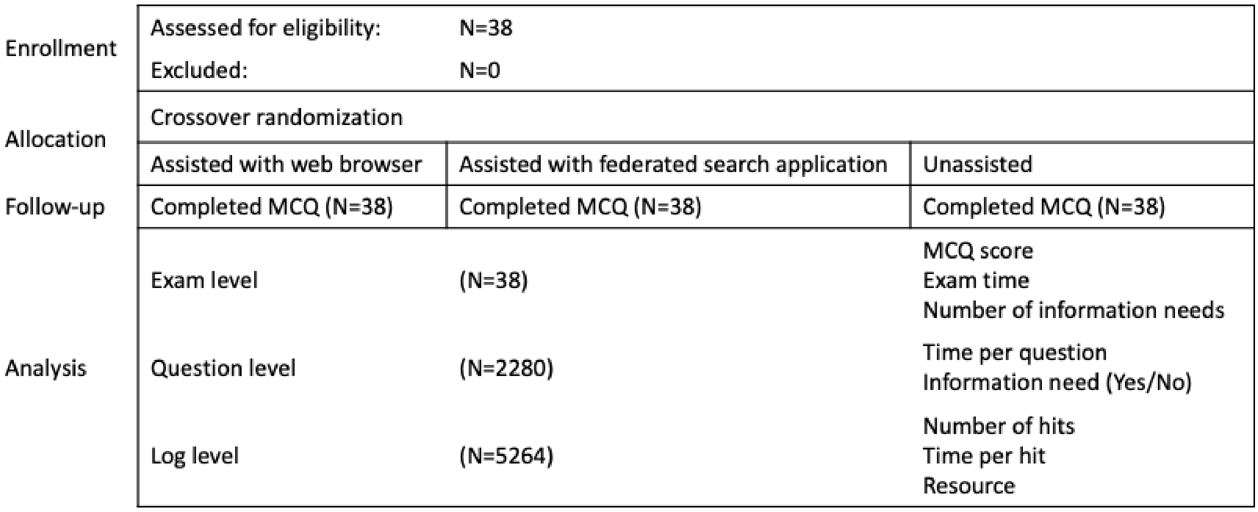
Participant Characteristics

A total of 38 physiotherapists fully participated in the study, completing an examination under each of the 3 experimental conditions in random order using Vickers' block randomization [42]. The demographic characteristics of the included sample are presented in Table 1. All tests were conducted in the 6-month period from June 2019 to January 2020. All participants completed all 3 conditions within a 1-month (4 week) period of enrolling in the study. A CONSORT diagram is presented in Figure 3 [43].

Table 1. Characteristics of participants.

Characteristics	Values
Male, n	19
Female, n	19
Age (years), mean (95% CI)	28.6 (27.5-29.7)
Number of years practicing, mean (95% CI)	5.4 (4.4-6.3)
Number of individual patient encounters per week, mean (95% CI)	26.1 (22.5-29.7)

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) diagram of the study design. Note that information needs were considered in the assisted conditions only. MCQ: multiple-choice questionnaire.



Enrollment	Assessed for eligibility: Excluded:	N=38 N=0	
Allocation	Crossover randomization		
	Assisted with web browser	Assisted with federated search application	Unassisted
Follow-up	Completed MCQ (N=38)	Completed MCQ (N=38)	Completed MCQ (N=38)
Analysis	Exam level	(N=38)	MCQ score Exam time Number of information needs
	Question level	(N=2280)	Time per question Information need (Yes/No)
	Log level	(N=5264)	Number of hits Time per hit Resource

Examination-Level Analysis

The examination-level GEE estimated a main effect for condition (*P*<.001). Examination scores differed significantly between the unassisted condition (mean examination score 47.11%, SD 12.7) and both assisted conditions (mean federated search condition 57.5%, SD 12.8, *P*=.001; web search condition 63.3%, SD 12%, *P*=.02). The mean examination score and examination duration for the 3 experimental conditions, and the mean number of information needs for the assisted conditions are presented in Table 2.

Table 2. Results of the examination-level analysis for the 3 experimental conditions including the mean examination score, duration, and number of information needs for the assisted conditions (with 95% CI).

Exam parameter	Number of information needs, mean (95% CI)	Time (min), mean (95% CI)	Score (%), mean (95% CI)
Assisted			
Federated search	11 (10-12)	34 (29-39)	58 (53-62)
Web search	11 (10-12)	27 (23-31)	63 (59-67)
Unassisted	N/A[a]	12 (11-14)	47 (43-51)

[a] N/A: not applicable.

Question-Level Analysis

Participants experienced 845 information needs out of a total of 1520 questions, with a rate of 55.6% (assisted conditions only). The question-level GEE was used to further investigate the relationship between individual information needs and answers (correct and incorrect). This GEE estimated a main effect for condition (*P*<.001) and the presence of an information need (*P*=.009). An analysis of the parameter estimates associated with these main effects revealed that participants were more likely to answer correctly in both the federated search condition (*P*=.003) and the web browser condition (*P*<.001) compared with the unassisted condition.

Despite the observation of a main effect for the presence of an information need, there were no significant differences based on the parameter estimates at the level of *P*<.05 in the rate of answering individual questions correctly in the presence or absence of an information need (*P*=.06).

The average rate of answering correctly for each condition in the absence or presence of an information need is presented in Table 3.

Table 3. Mean rate of answering correctly (with 95% CI) for each condition in the absence or presence of an information need.

Condition	Rate of answering correctly (%), mean (95% CI)		
	Federated search condition	Web search condition	Unassisted condition
Yes	62.1 (57.4-66.7)	65.2 (60.6-69.7)	N/A[a]
No	72.5 (67.6-77.2)	65.1 (60-70.2)	43.7 (40.1-47.2)

[a] N/A: not applicable.

Log Analysis

The 845 information needs corresponded to 1987 web logs and 3277 app logs. This data set of app and web logs comprised all participants' information-seeking sessions, which themselves included multiple information-seeking events (or information trails) split into primary, intermediate, and terminal periods.

Table 4 presents the most popular resources used during the primary, intermediate, and terminal periods in the web and app log conditions.

In the web-assisted condition, participants spent a total of 00:45:25 min (383 hits) in the primary period, 04:14:55 hours in the intermediate period (1356 hits), and 06:26:43 hours in the terminal period (248 hits). This corresponded with a median time per resource of 1 second in the primary period, 4 seconds per resource in the intermediate period, and 86 seconds per resource in the terminal period.

In the federated search condition, participants spent a total of 01:47:21 hours (486 hits) in the primary period, 08:42:24 hours in the intermediate period (2076 hits), and 04:24:06 hours in the terminal period (668 hits). This corresponded with a median time per resource of 3 seconds in the primary period, 4 seconds per resource in the intermediate period, and 5 seconds per resource in the terminal period.

The log GEE estimated a main effect for condition and for the period of the information trail (*P*<.001). There was also a significant main effect for the interaction between condition and period (*P*<.001). An analysis of the parameter estimates associated with these main effects revealed that participants spent less time per hit in the primary and intermediate periods for the web search condition compared with the federated search condition, and more time in the terminal period in the web search condition than in the federated search condition. The median and mean values for each condition stratified by search period are presented in Table 5.

Table 4. Most popular resources used during the primary, intermediate, and terminal periods for the web and federated search conditions.

Period, search condition, and resource	Hits, n	Cumulative (%)	Time (h:min:s)	Cumulative (%)
Primary				
Web browser				
Google Search	359	94	00:39:55	88
PubMed-NCBI[a]	9	96	00:02:35	94
Physio-pedia	9	98	00:01:32	97
ResearchGate	1	99	00:00:54	99
Google Scholar	2	99	00:00:15	99
Ovid	2	100	00:00:09	100
Federated search				
Wikipedia	304	62	00:53:35	50
PubMed	131	88	00:36:32	84
YouTube	36	95	00:12:13	95
Cochrane	23	100	00:05:03	100
Intermediate				
Web browser				
Google Search	858	63	01:33:50	37
PubMed-NCBI	141	74	00:49:23	56
Wikipedia	57	78	00:19:38	64
Physio-pedia	48	81	00:18:57	71
Google Scholar	47	85	00:10:07	75
British Journal of Sports Medicine	22	87	00:01:13	76
ResearchGate	14	88	00:04:25	77
Mayo Clinic	9	88	00:03:22	79
Sci-hub	8	89	00:00:56	79
Cochrane library	8	89	00:01:54	80
Federated search				
PubMed	937	46	03:56:48	46
Wikipedia	848	87	03:20:38	85
Cochrane	133	94	00:25:31	90
YouTube	129	100	00:51:18	100
Terminal				
Web browser				
Google Search	56	23	01:33:49	24
PubMed-NCBI	45	41	01:20:12	45
Physio-pedia	33	54	00:53:53	59
Wikipedia	26	65	00:35:35	68
British Journal of Sports Medicine	12	69	00:18:05	73
Mayo clinic	5	71	00:05:17	74
ResearchGate	5	73	00:01:27	75
Teachmeanatomy	4	75	00:06:19	76
Google Scholar	3	76	00:03:04	77
Mananatomy	3	77	00:05:16	78

Period, search condition, and resource	Hits, n	Cumulative (%)	Time (h:min:s)	Cumulative (%)
Federated search				
Wikipedia	366	55	01:47:48	42
PubMed	205	85	01:22:54	75
YouTube	57	94	00:50:49	94
Cochrane	42	100	00:14:04	100

[a] NCBI: National Center for Biotechnology Information.

Table 5. Median and mean values of time spent per resource for the web and federated condition stratified by search period.

Search period	Primary (second)	Intermediate (second)	Tertiary (second)
Federated search			
Median (IQR)	3 (11)	4 (12)	5 (26)
Mean (95% CI)	13.3 (10.6-15.9)	15.1 (13.6-16.6)	24 (20.3-27.2)
Web			
Median (IQR)	1 (7)	4 (14)	86 (55)
Mean (95% CI)	7.1 (5.3-9)	11.3 (10.4-12.2)	169.8 (40-299.5)

Discussion

Principal Findings

Frequency of Information Needs

In this study, we sought to confirm a series of hypotheses. The first concerned the frequency with which physiotherapists experience information needs. Information needs were specifically defined: they were induced during a theoretical examination and anchored to participants' access or nonaccess of digital information resources in the 2 "assisted" experimental conditions. Under this definition, participants experienced an information need in 55.59% (845/1520) of theoretical examination questions. In clinical environments, previous research has shown that medical physicians report experiencing an information need with varying frequencies. In the study by Covell et al [5], medical physicians had an information need in 67% of their patient encounters, whereas Sackett and Straus [44] observed 98 questions during the care of 166 hospitalized patients in a 30-day period (a rate of 59%). Among family care physicians, Ely et al [45] observed information needs at a rate of 3.2 (or 32%) for every 10 patient visits, while, more recently, Izcovich et al [46] documented 1.2 questions per patient encounter. It is important to note that by anchoring the existence of a need to information-seeking behavior, we could not evaluate *information needs* in the unassisted condition.

Regardless of how information needs are defined, timely translation of research evidence to the care pathway is a policy priority of many health research systems [47], yet many obstacles block the channels by which such translation occurs [48]. Ultimately, these obstacles impede effective evidence translation [49].

The Use of Web-Based Resources to Fulfill Information Needs

Barriers to obtaining information include time [40,50], accessibility [40,50,51], and limited personal skills [40], yet web-connected digital technologies could overcome these barriers and enable clinicians to fulfill their information needs at the point of care [1]. This contributed to our second confirmatory hypothesis that digital information resources accessed via the web could be used to fulfill information needs, thus improving the overall MCQ examination score. In agreement with this hypothesis, participants exhibited a mean improvement of 10% (for the federated search condition) and 16% (for the web condition) in the overall examination score compared with the unassisted condition. Our analysis at the resolution of individual questions further revealed that participants were more likely to answer correctly in either of the assisted conditions but that there were no differences in the rate of answering correctly in the presence or absence of an information need. Specifically, participants answered 63.7% (538/845) of questions correctly in the presence of an information need, compared with 68.7% (464/675) of questions in the absence of an information need. There was no statistically significant difference between these rates, suggesting that the availability of assistance improved participants' likelihood of answering a question correctly at a similar rate to that when they were confident of knowing the answer (and therefore did not seek assistance, even if it was available) or to a level where they were sufficiently confident to guess despite the expectation that they would be marked negatively for doing so.

Previous researchers have sought to affect knowledge acquisition as assessed via MCQ examinations among health care professionals with seminars [52], tutorials [53], and course modules [54]. Course materials are generally developed using the frameworks of evidence-based medicine [55,56]. However, this does not reflect real-world practice, as health care professionals [57,58] autonomously use search engines and nonauthoritative, community‐built content to fulfill their information needs, prioritizing efficiency, familiarity, accessibility, and organization with “just the right” volume, variety, and scope to fulfill their needs [40]. The apparent dissonance between the types of resources that academic staff and researchers encourage health care professionals to use [55,56], those they report using [59,60], and those they actually use [61] informed our hypothesis-generating research questions. Through the log file analysis, we evaluated the digital information resources used by our cohort of physiotherapists in an unconfined web environment and those accessed via a federated search engine that was developed for the purposes of this study.

Reflecting the available body of observational research evaluating the browsing behaviors of the general population for health-related information [62,63], we observed a preference mainly for search engines and nonauthoritative, community‐built content at all stages of the information-seeking journey. Specifically, in the web-assisted condition, Google was the most popular resource used at the start of the information-seeking journey (359/383, 93.7% of total hits; 00:39:55 of 00:45:25, 88% of the total time). The variety of resources then increased in the “intermediate” period, yet Google was still the most popular (858/1351, 63.51% of total hits; 01:33:50 of 04:13:03, 37% of total time), followed by PubMed (141/1351, 10.44% of total hits; 00:49:23 of 04:13:03, 20% of total time) and Wikipedia (57/1351, 4.22% of total hits; 00:19:38 of 04:13:03, 8% of total time). The terminal period contained the last resources used by participants before their submission of an answer in the MCQ. Therefore, these resources were assumed to be principally responsible for fulfilling participants’ information needs, an assumption reflected in the relative amount of time spent in this period (median duration of 86 seconds), compared with the primary (median duration of 1 second) and intermediate (median duration of 4 seconds) periods. In the terminal period, Google searches again accounted for the largest number of hits with 22.5% (56/248) of the total, which corresponded to 01:33:49, or 25% of the total duration (06:18:13), followed by PubMed (45/248, 18.1% of total hits; 01:20:12 of 06:18:13, 21% of total duration), Physio-pedia (33/248, 13.3% of total hits; 00:53:53 of 06:18:13, 14% of total duration), and Wikipedia (26/248, 10.4% of total hits; 00:35:35 of 06:18:13, 9% of total duration). That Google, PubMed, and Wikipedia were the most popular resources among a cohort of physiotherapists is in agreement with a small body of prior research in other health care professional groups [61]. However, that Google in particular is heavily relied upon as both a directory to other resources and as a resource itself is potentially problematic, owing to its propensity to display results in a “filter bubble” [64]. Indeed, Google was the only search engine used by participants in this study, and its tendency to display previews of the information contained in individual websites potentially bypasses the need to visit these websites and to properly apprise the evidence [65]. The impact of this in the context of health care delivery has yet to be formally evaluated, but the available body of research suggests that the use of Google may not always be conducive to acquiring valid and reliable health-related information [66,67].

The federated search engine in this study was designed as an alternative mode of assistance, bypassing the need for traditional web search filter bubbles and improving the efficiency of navigating specific academic databases. Unsurprisingly, due to the constraints on the resources that were included in the federated search, a less diverse array of information resources were observed in the federated search condition. Specifically, Wikipedia accounted for the greatest number of hits in the primary period (304/494, 61.5% of hits; 00:53:35 of 01:47:23, 50% of time), PubMed accounted for the greatest number of hits in the intermediate period (937 of 2047 hits and 03:56:48 of 08:34:14, or 45.77% of both total hits and time), and Wikipedia accounted for the greatest number of hits in the terminal period (366/670, 54.6% of total hits; 01:47:48 of 04:15:35, 42% time), suggesting that participants sought out these resources despite the differences in their mode of delivery. Although the difference was not statistically significant, participants spent longer doing the examination in the federated search condition (34 min vs 27 min in the web condition). However, participants did spend significantly less time in the terminal period (a median duration of 5 seconds, compared with 86 seconds in the web condition) and clicked on a greater number of resources (there were 1290, or 65% more hits in the federated search condition compared with the web condition). In summary, these findings may suggest that the federated search was less effective in finding relevant and useful information, required more effort to locate that information, or that participants used the primary, intermediate, and terminal resources together to fulfill their information needs, rather than reaching an information *end-point* through a terminal resource as in the web condition. In relation to our second hypothesis-generating research question, that there was no significant difference in the examination scores between the 2 assisted conditions, and no significant difference in the rate with which questions were answered correctly or incorrectly in the presence or absence of an information need, would suggest the latter conclusions to be more likely.

Limitations

Despite the methodological strengths of the crossover experimental design and the insights garnered via the use of web log analysis to determine what digital information resources were used to fulfill information needs and the novelty of evaluating these outcomes among physiotherapists, this study is not without limitations. First, an MCQ was used as a surrogate stimulus for information needs and may not be an accurate representation of the clinical setting. The validity of the MCQ questions as a measure of physiotherapy knowledge and by extension, health care delivery, was assumed on the basis that they were from the examinations of recognized accreditation bodies. It must be acknowledged that although these questions are used to assess physiotherapists’ capacity to practice, they may not actually represent the information needs of the authentic clinical encounter. Second, this study does not address an enduring question as to whether the fulfillment of clinical information needs alters treatment practices and patients’ outcomes; it has always been assumed, but never proven [46], that the translation of the latest high-quality research evidence to the care pathway optimizes clinical practice behaviors and improves patients' outcomes. Finally, although the use of a federated search portal system provides a useful comparator for unconstrained web browsing in evidence search and retrieval patterns and their effect on MCQ examination scores, the results presented in this study for this portal system (ie, SciScanner) are unique to it and have limited external applicability for other portal systems.

Conclusions

On the basis of an MCQ examination protocol, we identified that physiotherapists experienced an information need in 55.59% (845/1520) of theoretical questions and, when they were provided with access to digital information resources accessed via the web or a federated search software app, the fulfillment of these needs was associated with improved rates of answering examination questions correctly. The physiotherapists in this study exhibited a high preference for Google as both a directory and a resource, with Wikipedia and PubMed being the next most popular resources. The implications of relying heavily on Google as a search and retrieval mechanism for health-related information warrants further investigation, whereas the emulation of these findings in an authentic clinical setting would be an important research pursuit in the future.

Acknowledgments

This research was cofunded by the European Regional Development Fund under Ireland's European Structural and Investment Funds Programmes 2014-2020 under the *SciScanner* project title and the Health Research Board (ARPP-A-2018-002).

Conflicts of Interest

The authorship team (PM, AJ, and CD) was responsible for the development of the federated search portal app. The authors have no conflicts of interest to declare as the portal system accesses publicly available digital information resources unrelated to this group.

References

1. Mobasheri MH, King D, Johnston M, Gautama S, Purkayastha S, Darzi A. The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study. BMJ Innov 2015 Oct 7;1(4):174-181. [doi: 10.1136/bmjinnov-2015-000062]
2. Campbell R, Ash J. An evaluation of five bedside information products using a user-centered, task-oriented approach. J Med Libr Assoc 2006 Oct;94(4):435-41, e206 [FREE Full text] [Medline: 17082836]
3. Dicenso A, Bayley L, Haynes RB. Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. Evid Based Nurs 2009 Oct;12(4):99-101. [doi: 10.1136/ebn.12.4.99-b] [Medline: 19779069]
4. Pettigrew K, Fidel R, Bruce H. Conceptual frameworks in information behavior. Annual Review of information Science and Technology (ARIST) 2001;35:43-78 [FREE Full text]
5. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? Ann Intern Med 1985 Oct;103(4):596-599. [doi: 10.7326/0003-4819-103-4-596] [Medline: 4037559]
6. Case DO. Looking for information: a survey of research on information seeking, needs, and behavior. In: Library Management. Bingley, UK: Emerald Group Publishing; 2012.
7. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J 2007 Jun;24(2):78-94 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00713.x] [Medline: 17584211]
8. Daei A, Soleymani MR, Ashrafi-Rizi H, Zargham-Boroujeni A, Kelishadi R. Clinical information seeking behavior of physicians: a systematic review. Int J Med Inform 2020 Jul;139:104144. [doi: 10.1016/j.ijmedinf.2020.104144] [Medline: 32334400]
9. Osheroff JA, Forsythe DE, Buchanan BG, Bankowitz RA, Blumenfeld BH, Miller RA. Physicians' information needs: analysis of questions posed during clinical teaching. Ann Intern Med 1991 Apr 01;114(7):576-581. [doi: 10.7326/0003-4819-114-7-576] [Medline: 2001091]
10. Segev E, Sharon AJ. Temporal patterns of scientific information-seeking. Public Underst Sci 2016 May 20;26(8):969-985. [doi: 10.1177/0963662516648565]
11. Ahamad A, Wallner P, Salenius S, Ross R, Fernandez E. Information needs expressed during patient-oriented oncology consultations: quantity, variation, and barriers. J Cancer Educ 2019 Jun;34(3):488-497. [doi: 10.1007/s13187-018-1329-5] [Medline: 29435742]
12. Swennen M, van der Heijden GJ, Boeije HR, van Rheenen N, Verheul FJ, van der Graaf Y, et al. Doctors' perceptions and use of evidence-based medicine: a systematic review and thematic synthesis of qualitative studies. Acad Med 2013 Sep;88(9):1384-1396. [doi: 10.1097/ACM.0b013e31829ed3cc] [Medline: 23887011]
13. Currie L, Graham M, Allen M, Bakken S, Patel V, Cimino JJ. Clinical information needs in context: an observational study of clinicians while using a clinical information system. AMIA Annu Symp Proc 2003:190-194 [FREE Full text] [Medline: 14728160]

14. McKibbon KA, Fridsma DB. Effectiveness of clinician-selected electronic information resources for answering primary care physicians' information needs. J Am Med Inform Assoc 2006;13(6):653-659 [FREE Full text] [doi: 10.1197/jamia.M2087] [Medline: 16929042]
15. Borycki E, Lemieux-Charles L, Nagle L, Eysenbach G. Novice nurse information needs in paper and hybrid electronic-paper environments: a qualitative analysis. Stud Health Technol Inform 2009;150:913-917. [Medline: 19745445]
16. Rozic-Hristovsk A, Hristovski D, Todorovski L. Users' information-seeking behavior on a medical library website. J Med Libr Assoc 2002 Apr;90(2):210-217 [FREE Full text] [Medline: 11999179]
17. Bracke P. Web usage mining at an academic health sciences library: an exploratory study. J Med Libr Assoc 2004 Oct;92(4):421-428 [FREE Full text] [Medline: 15494757]
18. Meats E, Brassey J, Heneghan C, Glasziou P. Using the Turning Research Into Practice (TRIP) database: how do clinicians really search? J Med Libr Assoc 2007 Apr;95(2):156-163 [FREE Full text] [doi: 10.3163/1536-5050.95.2.156] [Medline: 17443248]
19. Lamprecht D, Lerman K, Helic D, Strohmaier M. How the structure of Wikipedia articles influences user navigation. New Rev Hypermedia Multimed 2017 Jan 2;23(1):29-50 [FREE Full text] [doi: 10.1080/13614568.2016.1179798] [Medline: 28670171]
20. Lamprecht D, Strohmaier M, Helic D, Nyulas C, Tudorache T, Noy NF, et al. Using ontologies to model human navigation behavior in information networks: a study based on Wikipedia. SW 2015 Aug 7;6(4):403-422. [doi: 10.3233/sw-140143]
21. Hersh W, Crabtree MK, Hickam DH, Sacherek L, Friedman CP, Tidmarsh P, et al. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. J Am Med Inform Assoc 2002;9(3):283-293 [FREE Full text] [doi: 10.1197/jamia.m0996] [Medline: 11971889]
22. Yoo I, Mosa AS. Analysis of pubmed user sessions using a full-day pubmed query log: a comparison of experienced and nonexperienced pubmed users. JMIR Med Inform 2015 Jul 2;3(3):e25 [FREE Full text] [doi: 10.2196/medinform.3740] [Medline: 26139516]
23. Maggio LA, Steinberg RM, Moorhead L, O'Brien B, Willinsky J. Access of primary and secondary literature by health personnel in an academic health center: implications for open access. J Med Libr Assoc 2013 Jul;101(3):205-212 [FREE Full text] [doi: 10.3163/1536-5050.101.3.010] [Medline: 23930091]
24. Herskovic JR, Tanaka LY, Hersh W, Bernstam EV. A day in the life of PubMed: analysis of a typical day's query log. J Am Med Inform Assoc 2007 Mar 1;14(2):212-220. [doi: 10.1197/jamia.m2191]
25. Islamaj Dogan R, Murray GC, Névéol A, Lu Z. Understanding PubMed user search behavior through log analysis. Database (Oxford) 2009;2009:bap018 [FREE Full text] [doi: 10.1093/database/bap018] [Medline: 20157491]
26. Scaffidi MA, Khan R, Wang C, Keren D, Tsui C, Garg A, et al. Comparison of the impact of Wikipedia, UpToDate, and a digital textbook on short-term knowledge acquisition among medical students: randomized controlled trial of three web-based resources. JMIR Med Educ 2017 Oct 31;3(2):e20 [FREE Full text] [doi: 10.2196/mededu.8188] [Medline: 29089291]
27. Kloda LA, Bartlett JC. A characterization of clinical questions asked by rehabilitation therapists. J Med Libr Assoc 2014 Apr;102(2):69-77 [FREE Full text] [doi: 10.3163/1536-5050.102.2.002] [Medline: 24860260]
28. Joorabchi A, Doherty C, Dawson J. 'WP2Cochrane', a tool linking Wikipedia to the Cochrane Library: results of a bibliometric analysis evaluating article quality and importance. Health Informatics J 2020 Sep;26(3):1881-1897 [FREE Full text] [doi: 10.1177/1460458219892711] [Medline: 31868082]
29. Physiotherapy Competency Exam. Canadian Alliance of Physiotherapy Regulators. 2019 Jan 01. URL: https://www.alliancept.org [accessed 2018-12-01] [WebCite Cache ID https://www.alliancept.org]
30. National Physical Therapy Exam. National Physical Therapy Association. URL: https://www.apta.org/Default.aspx [accessed 2018-12-01] [WebCite Cache ID https://www.apta.org/Default.aspx]
31. Fontinberry B. Q&A Review for the Physical Therapist Assistant Board Examination. Missouri, USA: Saunders; 2010.
32. Tyson A. NPTE Study Guide 2017: Practice Questions for the National Physical Therapy Examination. -: -; 2017.
33. ProProfs QuizMaker. ProProfs. 2019. URL: https://www.proprofs.com [accessed 2018-12-01]
34. Doherty C, Joorabchi P, Megyesi P. SciScanner: Search Science. 2020. URL: www.sciscanner.com [accessed 2019-12-01]
35. Alkhateeb F, Clauson KA, Latif DA. Pharmacist use of social media. Int J Pharm Pract 2011 Apr;19(2):140-142. [doi: 10.1111/j.2042-7174.2010.00087.x] [Medline: 21385246]
36. Egle JP, Smeenge DM, Kassem KM, Mittal VK. The internet school of medicine: use of electronic resources by medical trainees and the reliability of those resources. J Surg Educ 2015;72(2):316-320. [doi: 10.1016/j.jsurg.2014.08.005] [Medline: 25487347]
37. Allahwala UK, Nadkarni A, Sebaratnam DF. Wikipedia use amongst medical students – new insights into the digital revolution. Medical Teacher 2012 Nov 8;35(4):337. [doi: 10.3109/0142159x.2012.737064]
38. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011 Jan 31;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]
39. Cochrane and Wikipedia: Working Together to Improve Access to Health Evidence. Cochrane. 2017. URL: https://www.cochrane.org/news/cochrane-and-wikipedia-working-together-improve-access-health-evidence [accessed 2018-12-01]

40. Aakre C, Maggio LA, Fiol GD, Cook DA. Barriers and facilitators to clinical information seeking: a systematic review. J Am Med Inform Assoc 2019 Oct 1;26(10):1129-1140. [doi: 10.1093/jamia/ocz065] [Medline: 31127830]
41. -. PubMed clinical queries. Pediatric Neurology 2005 May;32(5):340. [doi: 10.1016/j.pediatrneurol.2005.03.011]
42. Vickers AJ. How to randomize. J Soc Integr Oncol 2006;4(4):194-198 [FREE Full text] [doi: 10.2310/7200.2006.023] [Medline: 17022927]
43. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
44. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the 'evidence cart'. J Am Med Assoc 1998 Oct 21;280(15):1336-1338. [doi: 10.1001/jama.280.15.1336] [Medline: 9794314]
45. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, et al. Analysis of questions asked by family doctors regarding patient care. Br Med J 1999 Aug 7;319(7206):358-361 [FREE Full text] [doi: 10.1136/bmj.319.7206.358] [Medline: 10435959]
46. Izcovich A, Malla CG, Diaz MM, Manzotti M, Catalano HN. Impact of facilitating physician access to relevant medical literature on outcomes of hospitalised internal medicine patients: a randomised controlled trial. Evid Based Med 2011 Oct;16(5):131-135. [doi: 10.1136/ebmed-2011-100117] [Medline: 21949275]
47. Westfall JM, Mold J, Fagnan L. Practice-based research—'Blue Highways' on the NIH Roadmap. J Am Med Assoc 2007 Jan 24;297(4):403. [doi: 10.1001/jama.297.4.403]
48. Sung NS, Crowley WF, Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. J Am Med Assoc 2003 Mar 12;289(10):1278-1287. [doi: 10.1001/jama.289.10.1278] [Medline: 12633190]
49. O'Connor PJ, Amundson G, Christianson J. Performance failure of an evidence-based upper respiratory infection clinical guideline. J Fam Pract 1999 Sep;48(9):690-697. [Medline: 10498075]
50. Dysart AM, Tomlin GS. Factors related to evidence-based practice among US occupational therapy clinicians. Am J Occup Ther 2002;56(3):275-284. [doi: 10.5014/ajot.56.3.275] [Medline: 12058516]
51. Hall EF. Physical therapists in private practice: information sources and information needs. Bull Med Libr Assoc 1995 Apr;83(2):196-201 [FREE Full text] [Medline: 7599585]
52. Amsallem E, Kasparian C, Cucherat M, Chabaud S, Haugh M, Boissel J, et al. Evaluation of two evidence-based knowledge transfer interventions for physicians. A cluster randomized controlled factorial design trial: the CardioDAS Study. Fundam Clin Pharmacol 2007 Dec;21(6):631-641. [doi: 10.1111/j.1472-8206.2007.00513.x] [Medline: 18034664]
53. Considine J, Brennan D. Effect of an evidence-based paediatric fever education program on emergency nurses' knowledge. Accid Emerg Nurs 2007 Jan;15(1):10-19. [doi: 10.1016/j.aaen.2006.11.005] [Medline: 17218101]
54. Widyahening IS, Findyartini A, Ranakusuma RW, Dewiasty E, Harimurti K. Evaluation of the role of near-peer teaching in critical appraisal skills learning: a randomized crossover trial. Int J Med Educ 2019 Jan 25;10:9-15. [doi: 10.5116/ijme.5c39.b55b] [Medline: 30685751]
55. Ardern CL, Dupont G, Impellizzeri FM, O'Driscoll G, Reurink G, Lewin C, et al. Unravelling confusion in sports medicine and sports science practice: a systematic approach to using the best of research and practice-based evidence to make a quality decision. Br J Sports Med 2019 Jan;53(1):50-56. [doi: 10.1136/bjsports-2016-097239] [Medline: 28818957]
56. Crowell MS, Tragord BS, Taylor AL, Deyle GD. Integration of critically appraised topics into evidence-based physical therapist practice. J Orthop Sports Phys Ther 2012 Oct;42(10):870-879. [doi: 10.2519/jospt.2012.4265] [Medline: 22814199]
57. Mikalef P, Kourouthanassis PE, Pateli AG. Online information search behaviour of physicians. Health Info Libr J 2017 Mar;34(1):58-73. [doi: 10.1111/hir.12170] [Medline: 28168831]
58. Batt-Rawden S, Flickinger T, Weiner J, Cheston C, Chisolm M. The role of social media in clinical excellence. Clin Teach 2014 Jun 10;11(4):264-269. [doi: 10.1111/tct.12129]
59. de Leo G, LeRouge C, Ceriani C, Niederman F. Websites most frequently used by physician for gathering medical information. AMIA Annu Symp Proc 2006:902 [FREE Full text] [Medline: 17238521]
60. Shariff SZ, Bejaimal SA, Sontrop JM, Iansavichus AV, Weir MA, Haynes RB, et al. Searching for medical information online: a survey of Canadian nephrologists. JN 2011 Feb 23;24(6):723-732. [doi: 10.5301/jn.2011.6373]
61. Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 2009 Oct;78(10):645-655. [doi: 10.1016/j.ijmedinf.2009.04.008] [Medline: 19501017]
62. Lee K, Hoti K, Hughes JD, Emmerton L. Dr google is here to stay but health care professionals are still valued: an analysis of health care consumers' internet navigation support preferences. J Med Internet Res 2017 Jun 14;19(6):e210. [doi: 10.2196/jmir.7489] [Medline: 28615156]
63. Top 500. Alexa Company. 2019. URL: https://www.alexa.com [accessed 2019-02-27]
64. Measuring the 'Filter Bubble': How Google is Influencing What You Click. DuckDuckGo. 2018. URL: https://spreadprivacy.com/google-filter-bubble-study/ [accessed 2019-12-01]
65. Search Console Help. Google. URL: https://support.google.com/webmasters/answer/7358659?hl=en [accessed 2018-12-01]

66. Kingsley K, Galbraith GM, Herring M, Stowers E, Stewart T, Kingsley KV. Why not just Google it? An assessment of information literacy skills in a biomedical science curriculum. BMC Med Educ 2011 Apr 25;11:17 [FREE Full text] [doi: 10.1186/1472-6920-11-17] [Medline: 21518448]
67. Pusz MD, Brietzke SE. How good is Google? The quality of otolaryngology information on the internet. Otolaryngol Head Neck Surg 2012 Sep;147(3):462-465. [doi: 10.1177/0194599812447733] [Medline: 22597577]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
CSV: comma-separated values
GEE: generalized estimating equation
MCQ: multiple-choice questionnaire
NPTE: National Physical Therapy Exam
PCE: Physiotherapy Competency Exam

Edited by G Eysenbach, R Kukafka; submitted 30.04.20; peer-reviewed by E Prakash, E Burner, T Gladman; comments to author 30.06.20; revised version received 06.07.20; accepted 11.11.20; published 17.12.20.

Please cite as:
Doherty C, Joorabchi A, Megyesi P, Flynn A, Caulfield B
Physiotherapists' Use of Web-Based Information Resources to Fulfill Their Information Needs During a Theoretical Examination: Randomized Crossover Trial
J Med Internet Res 2020;22(12):e19747
URL: http://www.jmir.org/2020/12/e19747/
doi:10.2196/19747
PMID:33331826

Original Paper

Psychometric Validation and Cultural Adaptation of the Simplified Chinese eHealth Literacy Scale: Cross-Sectional Study

Richard Huan Xu[1*], PhD; Lingming Zhou[2*], MSc; Sabrina Yujun Lu[3], MPH; Eliza Laiyi Wong[1], PhD; Jinghui Chang[2], MA; Dong Wang[2], PhD

[1]JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[2]School of Health management, Southern Medical University, Guangzhou, China
[3]Department of Sports Science and Physical Education, Chinese University of Hong Kong, Hong Kong, Hong Kong
[*]these authors contributed equally

Corresponding Author:
Dong Wang, PhD
School of Health management
Southern Medical University
No. 1023 Shatai nan Road
Guangzhou
China
Phone: 86 61648573
Email: dongw96@smu.edu.cn

Abstract

Background: The rapid proliferation of web-based information on health and health care has profoundly changed individuals' health-seeking behaviors, with individuals choosing the internet as their first source of information on their health conditions before seeking professional advice. However, barriers to the evaluation of people's eHealth literacy present some difficulties for decision makers with respect to encouraging and empowering patients to use web-based resources.

Objective: This study aims to examine the psychometric properties of a simplified Chinese version of the eHealth Literacy Scale (SC-eHEALS).

Methods: Data used for analysis were obtained from a cross-sectional multicenter survey. Confirmatory factor analysis (CFA) was used to examine the structure of the SC-eHEALS. Correlations between the SC-eHEALS and ICEpop capability measure for adults (ICECAP-A) items and overall health status were estimated to assess the convergent validity. Internal consistency reliability was confirmed using Cronbach alpha (α), McDonald omega (ω), and split-half reliability (λ). A general partial credit model was used to perform the item response theory (IRT) analysis. Item difficulty, discrimination, and fit were reported. Item-category characteristic curves (ICCs) and item and test information curves were used to graphically assess the validity and reliability based on the IRT analysis. Differential item functioning (DIF) was used to check for possible item bias on gender and age.

Results: A total of 574 respondents from 5 cities in China completed the SC-eHEALS. CFA confirmed that the one-factor model was acceptable. The internal consistency reliability was good, with α=0.96, ω=0.92, and λ=0.96. The item-total correlation coefficients ranged between 0.86 and 0.91. Items 8 and 4 showed the lowest and highest mean scores, respectively. The correlation coefficients between the SC-eHEALS and ICECAP-A items and overall health status were significant, but the strength was mild. The discrimination of SC-eHEALS items ranged between 2.63 and 5.42. ICCs indicated that the order of categories' thresholds for all items was as expected. In total, 70% of the information provided by SC-eHEALS was below the average level of the latent trait. DIF was found for item 6 on age.

Conclusions: The SC-eHEALS has been demonstrated to have good psychometric properties and can therefore be used to evaluate people's eHealth literacy in China.

(J Med Internet Res 2020;22(12):e18613) doi:10.2196/18613

KEYWORDS

electronic health literacy; eHEALS; psychometric property; classical test theory; item response theory; China

Introduction

Background

Currently, 4 billion people worldwide use the internet for various purposes every day, which increases their ability to search, digest, and use information in every aspect of their lives [1]. The health care field cannot escape from this trend. The rapid proliferation of web-based information about health and health care has significantly changed individuals' health-seeking behavior, with individuals choosing the internet as their first source of information regarding their health conditions before seeking professional advice [2,3].

Despite increased accessibility to relevant health information on the internet, searching, identifying, and using the *right* information is still challenging. Van der Vaart et al [4] found that easily identifiable web-based information may increase the discrepancies in health knowledge and choice of health care provider. Collecting web-based information is different from acquiring health-related information via traditional methods, such as hospital pamphlets, medical books, or professionals' advice; acquiring useful and accurate web-based information necessitates specific skills [3,5,6]. These skills include not only having professional knowledge about any given health conditions but also computer literacy, mobile phone literacy, and knowing how to navigate the internet [4,7]. Thus, eHealth literacy, including computer literacy and internet-related knowledge and skills, plays a key role in helping people search for web-based information and analyze, assess, and use that information to improve their health outcomes [8]. eHealth information offers little value if the intended users lack the skills to effectively use this type of information [9]. Measuring people's eHealth literacy could help policy makers develop guidelines, strategies, and interventions to provide health information through the internet to people regarding different needs, preferences, and expectations in ways that they can understand and use it [7].

In the last two decades, China has made remarkable progress in the development of internet networks and services. Currently, more than 800 million Chinese people actively use the internet every day [10]. To maximize the effectiveness of the internet in improving people's health, in 2018, the State Council of China announced an ambitious plan to develop an integral system to provide a broad spectrum of health care services through the internet across China in the next 10 years; this initiative intends to solve the long-term problems of *kanbing nan and kanbing gui* (too inaccessible and expensive to visit physicians) [11]. An instrument that comprehensively evaluates people's skills and literacy to understand and use web-based information and services is essential. In particular, given that China has a dual social urban-rural system, the gap between the cities and the countryside, including economic, educational, health care, and so on, was and continues to be tremendous. The provision of consumer-tailored eHealth information that meets their levels of eHealth literacy is the key to engaging them in making sound health decisions [9].

At present, few measures exist to specifically measure eHealth literacy, which has a different conceptual structure compared with traditional health literacy. In 2006, the eHealth literacy scale (eHEALS) was developed to focus on assessing the skills for finding, evaluating, and applying web-based health information to improve one's health outcomes [12]. It has been translated and validated in several countries and has been demonstrated to be a valuable instrument for assessing the effects of eHealth literacy on helping decision makers develop appropriate strategies to provide web-based information to improve people's health.

In China, several studies have reported the performance of eHEALS in different settings, and all of them showed some limitations. The generalizability of the findings from the study by Chang and Schultz [13] was questionable because they used a total of 352 patients from one selected hospital from a city in China, and their targeted populations were only those who reported having chronic conditions. For the study by Ma and Wu [14], the major limitation was that all the respondents included in that study were rural residents and were recruited from one district of a small city in Western China (the poorest part of China). By 2018, approximately 60% of the population in China lived in urban areas; therefore, assessing the psychometric properties of the eHEALS on only rural residents from one district would lead to ubiquitous selection and information bias [14]. Furthermore, the study by Zibrik et al [15] also reported the psychometric properties of the eHEALS; however, their study was not conducted in mainland China, and the targeted population was British Columbia's immigrant Chinese and Punjabi senior population. The study by Guo et al [16] seemed to be the first psychometric study of the Chinese eHEALS; however, 2 issues need to be addressed. First, their work was presented in Chinese rather than English, where the results are difficult for non-Chinese researchers to understand—this limits international or regional comparisons within Asian countries or between Chinese populations from different countries. Second, the targeted population of that study was 110 high school students recruited from Beijing (the most developed region in China). The selection and information bias could not be neglected. Moreover, except for the study by Ma and Wu [14], no previous studies used both classical test theory (CTT) and item response theory (IRT) to assess the validity of the Chinese eHEALS, and no study has adopted the differential item functioning (DIF) analysis to investigate the item variance of the scale. Considering the limitations summarized above, it is important to assess the performance of the Chinese version of eHEALS using more comprehensive methods and a better representative sample, which covers respondents from developed and underdeveloped areas, urban and rural residents, and across a wide age range.

Aims of the Study

This study aims to examine the psychometric properties of a simplified Chinese version of the eHEALS (SC-eHEALS) in the Chinese population based on a multicenter cross-sectional study.

Methods

Translation

The translation of the SC-eHEALS followed the international protocol [17]. Two translators, who were native Chinese speakers but were fluent in English, were invited to translate the eHEALS from English into simplified Chinese independently. The 2 transcripts were collected by the local research team for discussion and synthesis. Subsequently, a revised version was produced and sent to 2 professional translators for independent backward translation. The local research team examined the back translation against the original English version to identify any discrepancies, addressed the disputed items, and refined the translation, focusing on cultural adaptation until consensus was achieved by all the research team members. Then, a cognitive debriefing was carried out that involved 15 patients recruited from a hospital in Guangzhou. All the respondents were asked to comment on the response options and any word or phrasing that they found difficult to understand. Respondents were asked to describe in their own language what the word or phrasing meant to them. The results of the pilot study were discussed by the local research team, and then the final version of the SC-eHEALS was verified.

Study Design and Data Collection

Data used in this study were obtained from a multicenter cross-sectional survey that investigated patients' attitudes toward patient-centered care (PCC) in China from November 2019 to January 2020. Patients were recruited from the inpatient departments of 8 hospitals from 5 cities (Guangzhou, Shenzhen, Zhanjiang, Meizhou, and Shaoguan) of the Guangdong province (Southern China). All patients from the target hospitals during the appointed survey period were invited to participate in the survey. The inclusion criteria were as follows: (1) aged ≥18 years, (2) can read and speak Chinese, (3) no cognitive problems, and (4) able to provide informed consent. A team of investigators with rich experience in conducting face-to-face interviews was recruited to carry out the fieldwork. With the assistance of ward nurses, all respondents were asked to complete a structured questionnaire that included questions about demographic characteristics, socioeconomic status (SES), health conditions, well-being, health services use, lifestyle, and attitudes toward PCC. To conduct confirmatory factor analysis (CFA), the minimum sample size required was 300 [18,19]. For the IRT analysis, the preferred sample size required is 500 for analyzing a scale comprising polytomous items [20]. In total, the data of 574 patients who completed the SC-eHEALS questionnaire were elicited for our psychometric analysis. Informed consent was obtained from all respondents. The study protocol was approved by the institutional review board of the second affiliated hospital of Guangzhou Medical University (ethical approval ID: 2019-ks-28).

Instrument

eHEALS

eHEALS is used to measure consumers' combined knowledge, comfort, and perceived skills at finding, evaluating, and applying eHealth information to manage health problems [12]. It was developed based on a framework that consists of 6 dimensions to understand and use eHealth information [9]. The eHEALS has 8 items that are rated on a 5-point Likert scale ranging from *strongly disagree* to *strongly agree*. A one-factor structure was confirmed in the original study, and the reliability was acceptable with a Cronbach alpha of .88 [12]. The sum score of eHEALS ranges from 8 to 40, where a higher score indicates greater perceived eHealth literacy. The translation of eHEALS to SC-eHEALS has been discussed earlier.

The ICEpop CAPability Measure for Adults

The ICEpop capability measure for adults (ICECAP-A) is a generic and preference-based instrument that evaluates an individual's well-being [21]. It has 5 items (stability, attachment, autonomy, achievement, and enjoyment), and each item has 4 response options that range from fully capable to not capable.

Overall Health Status

The respondents' overall health status was evaluated using a visual analogue scale (VAS). They were presented with a scale ranging from 0 to 100, where 0 represents the worst health status and 100 represents the best health status they can imagine on the surveying day.

Statistical Analysis

Construct Validity

CFA was used to examine the structure of SC-eHEALS. The model fit was determined by 4 indicators, that is, root mean square error of approximation (RMSEA≤0.08, fair), the comparative fit index (CFI>0.95, good), the Tucker-Lewis index (TLI>0.95, good), and standardized root mean square residual (SRMR<0.08, acceptable) [22]. Moreover, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) were supplemented to compare the performance of the models, with smaller values indicating a better fit. Exploratory factor analysis (EFA) was used when the one-factor model reported in the original study was not supported by CFA.

Convergent Validity

Convergent validity was confirmed by evaluating the correlations between the SC-eHEALS and ICECAP-A items and the VAS item. Pearson correlation coefficient (r) was used to assess the strength of the correlation with r>0.3 identified as *moderate* [23].

Known-Group Validity

One-way analysis of variance (ANOVA) was used to assess the known-group validity of the SC-eHEALS. As individuals' levels of eHealth literacy were impossible to observe directly, on the basis of literature review, we hypothesized that the patients who were young, highly educated, or fully employed (representing good SES) would show a high degree of eHealth literacy.

Item Statistics and Internal Consistency Reliability

The mean and SD of the SC-eHEALS scores at both the item- and scale-level were reported as well as the ceiling and floor effects. The internal consistency reliability was estimated using Cronbach alpha (α>.7, acceptable), McDonald omega (ω>0.7, acceptable), and split-half reliability (λ>0.7, acceptable).

Item-total correlations (>0.7 acceptable) and alpha were presented if items were deleted [23].

IRT Analysis

Considering that there is no gold standard for model selection to perform IRT analysis, we adopted 3 models in this study, compared their performance, and reported the results of the best-fit model. The first model was the partial credit model (PCM), which is an extended version of the Rasch model. The second model was the general partial credit model (GPCM), which is a modified version of PCM and used to estimate 2 parameters in the analysis; and the last was the rating scale model, which is another Rasch family model that requires all items to have the same number of options. The details of the IRT models can be found in the study by DeMars [24]. The likelihood ratio test was used to compare the performance of the models. AIC and BIC were also used to assess the model fit. The results of model comparisons showed the GPCM outperformed the other 2 models [25]. For GPCM analysis, the (1) discrimination, (2) difficulty, and (3) item fit (the value of S-$\chi 2$) were calculated for each item, along with the item information curves (IICs) and test information curve (TIC) [26]. The item-category characteristic curves (ICCs) graphically presented the probability of the response to a given item in each category as the function of the latent trait (eHealth literacy).

Differential Item Functioning

DIF was used to check for the possible item bias caused by responses from different subgroups (gender, age, family registry, chronic conditions, and educational level) in the sample [27]. Pseudo McFadden R^2 was used to estimate the magnitude of the DIF. $R^2<0.13$ was deemed as *negligible*, 0.13 to 0.26 as *moderate*, and >0.26 as *large*. [28]

R (R foundation) was used for data analysis, and the *P* value was set at <.05. For IRT analysis, the Bonferroni correction was applied, and the *P* value was set at <.006 (.05/8).

Results

Respondents' Characteristics

In this study, more than half of the respondents were men (292/574, 50.9%), and the mean age was 45.58 years (SD 16.18). Almost half of them completed tertiary educational attainment (263/574, 45.8%), and 50.8% (231/574) were urban residents. Nearly half of the respondents reported living with chronic conditions (281/574, 49.1%) and only one fifth of the respondents reported that their disease was a severe threat to their lives (112/574, 19.8%; Table 1).

Table 1. Respondents' characteristics (n=574).

Characteristic	Value
Sex, n (%)	
Male	292 (50.9)
Female	281 (48.9)
Age (years), mean (SD)	45.58 (16.18)
Education, n (%)	
No or primary	93 (16.2)
Secondary or postsecondary	218 (37.9)
Tertiary or above	263 (45.8)
Family registry, n (%)	
Urban	291 (50.8)
Rural	282 (49.2)
Caregiver, n (%)	
No	418 (72.8)
Yes	156 (27.2)
Living status, n (%)	
Live alone	57 (9.9)
Live with family or others	517 (90.1)
Working status, n (%)	
Employed	394 (68.6)
Unemployed	180 (31.4)
Income level[a](CNY ¥ [US $]), n (%)	
≤900[b] (64.3)	112 (19.6)
901-1800 (64.4-128.6)	48 (8.4)
1801-2700 (128.6-192.9)	68 (11.9)
2701-3800 (192.9-271.4)	78 (13.7)
3801-6400 (271.4-457.1)	125 (21.9)
≥6401 (457.1)	140 (24.5)
BMI, n (%)	
<23	294 (51.7)
≥23	275 (48.3)
Chronic condition, n (%)	
Yes	281 (49.1)
No	291 (50.9)
Self-reported health condition, n (%)	
Severe threat to life	112 (19.8)
Moderate threat to life	112 (19.8)
Mild threat to life	137 (24.2)
No threat to life	206 (35.8)

[a] Disposable personal income per month.

[b] 100 CNY ¥ equals approximately 14 US $ (February 2020 exchange rate).

Construct Validity

The results of the CFA showed that the model fit of the one-factor model was not satisfactory with a low TLI and high RMSEA values (Table 2). We further examined items with high residual correlations and modification indices. On examining the wordings and polychoric correlation of items, we found that the nonrandom measurement error was caused by item redundancy. The fit of the revised model was improved after we modified the model and specified the error covariance between items 1 and 2, 3 and 4, and 7 and 8. A two-factor structure was suggested by the EFA (Multimedia Appendix 1). The CFA indicated that the performance of the revised two-factor model was not better than the revised one-factor model. After checking the item correlations and residuals, a revised two-factor model was confirmed, with RMSEA=0.08, SRMR=0.02, CFI=0.99, and TLI=0.96. The revised two-factor model outperformed the other models. The standardized factor loadings for the observed variables of all CFA models are presented in Multimedia Appendix 2.

Table 2. Confirmatory factor analysis of the simplified Chinese eHealth literacy scale.

Model	Model fit statistics							
	χ^2 (*df*)	*P* value	RMSEA[a] (95% CI)	CFI[b]	TLI[c]	SRMR[d]	AIC[e]	BIC[f]
One-factor	443.56 (20)	<.001	0.19 (0.18-0.21)	0.92	0.88	0.04	7905.29	7974.93
Revised one-factor	134.76 (17)	<.001	0.11 (0.09-0.12)	0.97	0.96	0.03	7602.49	7685.19
Two-factor	252.73 (19)	<.001	0.15 (0.13-0.16)	0.95	0.93	0.03	7716.46	7790.45
Revised two-factor	83.2 (17)	<.001	0.08 (0.06-0.1)	0.99	0.98	0.02	7550.93	7633.63

[a] RMSEA: root mean square error of approximation.

[b] CFI: comparative fit index.

[c] TLI: Tucker-Lewis index.

[d] SRMR: standardized root mean square residual.

[e] AIC: Akaike information criterion.

[f] BIC: Bayesian information criterion.

Convergent and Known-Group Validity

Table 3 shows the correlations between the SC-eHEALS and ICECAP-A items and overall health status. All the correlations of SC-eHEALS with the other measures were statistically significant, but the value of coefficients indicated a mild correlation. The results of ANOVA indicate that respondents who were young, highly educated, and fully employed tended to report a high level of eHealth literacy, which indicated that the known-group validity of the SC-eHEALS was satisfied (Table 4).

Table 3. Convergent validity of the simplified Chinese eHealth literacy scale.

Simplified Chinese eHealth literacy scale items[a]	ICECAP-A[b]										VAS[c]	
	Stability	*P* value	Attachment	*P* value	Enjoyment	*P* value	Achievement	*P* value	Autonomy	*P* value	Value	*P* value
eHEALS[d] 1	−0.15	<.001	−0.16	<.001	−0.12	.003	−0.11	.007	−0.17	<.001	0.12	.006
eHEALS 2	−0.13	.002	−0.14	.001	−0.12	.003	−0.13	.002	−0.12	.004	0.11	.007
eHEALS 3	−0.16	<.001	−0.19	<.001	−0.18	<.001	−0.18	<.001	−0.23	<.001	0.10	.02
eHEALS 4	−0.15	<.001	−0.14	<.001	−0.15	<.001	−0.11	.01	−0.20	<.001	0.10	.02
eHEALS 5	−0.17	<.001	−0.15	<.001	−0.15	<.001	−0.14	.001	−0.17	<.001	0.09	.05
eHEALS 6	−0.16	<.001	−0.20	<.001	−0.14	<.001	−0.14	.001	−0.19	<.001	0.11	.007
eHEALS 7	−0.18	<.001	−0.13	<.001	−0.17	<.001	−0.18	<.001	−0.15	<.001	0.17	<.001
eHEALS 8	−0.20	<.001	−0.17	<.001	−0.18	<.001	−0.16	<.001	−0.17	<.001	0.14	.001
Sum score	−0.18	<.001	−0.19	<.001	−0.16	<.001	−0.16	<.001	−0.20	<.001	0.14	.001

[a] For the eHealth literacy scale, a higher score indicates good eHealth.

[b] For ICECAP-A, a higher score indicates worse status of well-being.

[c] VAS: visual analogue scale of overall physical health.

[d] eHEALS: eHealth literacy scale.

Table 4. Known-group validity of the simplified Chinese eHealth literacy scale.

Characteristics	Mean (SD)	*P* value
Age (years)		.003
≤30	30.85 (6.28)	
31-60	29.34 (6.42)	
≥61	27.28 (7.91)	
Education		.02
No or primary	28.29 (7.99)	
Secondary or postsecondary	28.46 (7.05)	
Tertiary or above	30.27 (5.93)	
Working status		.003
Employed	29.95 (6.46)	
Unemployed	27.75 (7.22)	

Item Statistics and Internal Consistency Reliability

Table 5 shows that item 8 was rated as the most difficult item with a mean score of 3.49 (SD 1.01), whereas item 4 was rated as the easiest item with a mean score of 3.75 (SD 0.93). Items showed no floor effects, but some ceiling effects, which ranged from 16.9% to 19.69%. The reliability of the scale was good, as the Cronbach alpha was .96, McDonald omega was 0.92 and split-half reliability was 0.96. In addition, the value of alpha if an item was deleted was lower than that observed if the item was retained, and the item-total correlation coefficients ranged between 0.86 and 0.91. The SC-eHEALS scores stratified according to respondents' characteristics are presented in Multimedia Appendix 3.

Table 5. The response pattern, item statistics, and internal consistency reliability of the simplified Chinese eHealth literacy scale.

SC-eHEALS[a]	Response, n					Item statistics			Internal consistency reliability	
	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean (SD)	Floor, n (%)	Ceiling, n (%)	Alpha if item deleted[b]	Item-total correlation[b]
eHEALS[c] 1	18	36	150	267	103	3.70 (0.94)	18 (3.14)	103 (17.94)	.88	0.91
eHEALS 2	16	29	156	264	109	3.73 (0.92)	16 (2.79)	109 (18.99)	.88	0.91
eHEALS 3	14	39	141	278	102	3.72 (0.92)	14 (2.44)	102 (17.77)	.82	0.86
eHEALS 4	15	37	135	274	113	3.75 (0.93)	15(2.61)	113(19.69)	.86	0.89
eHEALS 5	17	49	160	242	106	3.65 (0.97)	17 (2.96)	106 (18.47)	.87	0.90
eHEALS 6	17	45	169	242	101	3.64 (0.96)	17 (2.96)	101 (17.60)	.86	0.89
eHEALS 7	20	47	195	208	104	3.57 (0.99)	20 (3.48)	104 (18.12)	.84	0.88
eHEALS 8	22	58	205	192	97	3.49 (1.01)	22 (3.83)	97 (16.90)	.81	0.86
Overall	N/A[d]	N/A	N/A	N/A	N/A	29.26 (6.78)[e]	8 (1.57)	40 (10.63)	N/A	N/A

[a] SC-eHEALS: simplified Chinese eHealth literacy scale.

[b] Cronbach alpha=.96; McDonald omega=.92; split-half reliability=.96.

[c] eHEALS: eHealth literacy scale.

[d] N/A: not applicable.

[e] The overall score of the eHEALS ranges between 0 and 40.

IRT Analysis

The results of the GPCM analysis are presented in Table 6. The discrimination indices of items ranged between 2.63 and 5.42, which indicated that the SC-eHEALS can distinguish between individuals with either low or high eHealth literacy, corresponding with the latent trait sensitivity. The item difficulty of threshold 1 and threshold 4 ranged from −2.03 to −1.79 and 0.84 to 1.04, respectively. The *P* value of S-χ^2 (Specific item-fit index) reflected that items 3 (*P*<.001), 4 (*P*=.003), and 8 (*P*<.001) might show misfit to some extent. Most of the information (70%) provided by the SC-eHEALS was between −6 and 0 on the latent trait scale (the comparisons between different IRT models are presented in Multimedia Appendix 4, and the misfit items of the SC-eHEALS are presented in Multimedia Appendix 5).

Table 6. The result of item response theory analysis of the simplified Chinese eHealth literacy scale.

SC-eHEALS[a]	Discrimination of item	Difficulty of item				Specific item-fit index, S-χ^2 (df)	*P* value[b]	Information (0-100[c])	
		T1[d]	T2[e]	T3[f]	T4[g]			−6 to 0	0 to 6
eHEALS 1	5.27	−1.79	−1.27	−0.41	0.90	19.041 (17)	.33	72.65	27.35
eHEALS 2	5.42	−1.85	−1.37	−0.42	0.84	24.171 (16)	.09	72.95	27.05
eHEALS 3	3.09	−2.03	−1.37	−0.51	1.01	61.57 (21)	<.001	71.62	28.38
eHEALS 4	4.17	−1.94	−1.31	−0.50	0.85	38.969 (18)	.003	72.95	27.05
eHEALS 5	4.52	−1.87	−1.17	−0.33	0.89	32.962 (19)	.05	70.73	29.27
eHEALS 6	3.83	−1.89	−1.24	−0.31	0.96	27.916 (20)	.11	69.67	30.33
eHEALS 7	3.24	−1.80	−1.26	−0.17	0.95	41.64 (24)	.01	66.58	33.42
eHEALS 8	2.63	−1.80	−1.20	−0.06	1.04	73.951 (25)	<.001	64.01	35.99
Overall	N/A[h]	N/A	N/A	N/A	N/A	N/A	N/A	70.70	29.30

[a] SC-eHEALS: simplified Chinese eHealth literacy scale.

[b] Items with *P*<.05/8=.0062 were flagged as potentially misfit.

[c] 0 means no information, 100 means full information.

[d] T1: threshold categories 1 and 2.

[e] T2: threshold categories 2 and 3.

[f] T3: threshold categories 3 and 4.

[g] T4: threshold categories 4 and 5.

[h] N/A: not applicable.

The ICCs and IICs for SC-eHEALS are graphically presented in Figures 1 and 2, respectively. The curves of the ICCs showed that the order of categories' thresholds for all the items was as expected, which meant that all categories were adequate in terms of placing a respondent on the scale. The distributions of the IICs were multimodal. The shapes of items 1 and 2 were the steepest and provided more information than the other items. The shapes of items 3 and 8 were the flattest, which means little information was provided by these items. TIC of the SC-eHEALS is presented in Multimedia Appendix 6.

Item-person map displays the location parameters of items as well as the distribution of person parameters along the same latent dimension. We found that the gap between threshold 3 and threshold 4 was larger than the other gaps; however, the discrepancy was diminished from item 1 to 8. In contrast, the gap between threshold 1 and threshold 2 was smaller than the others; however, the discrepancy increased from items 1 to 8. The distribution of person parameters showed some respondents located at the high end of the latent trait scale, which means they are very likely equipped with strong eHealth literacy (Figure 3).

Figure 1. Item-category characteristic curves for the simplified Chinese eHealth literacy scale. eHEALS: eHealth literacy scale; P: probability of option to be selected.

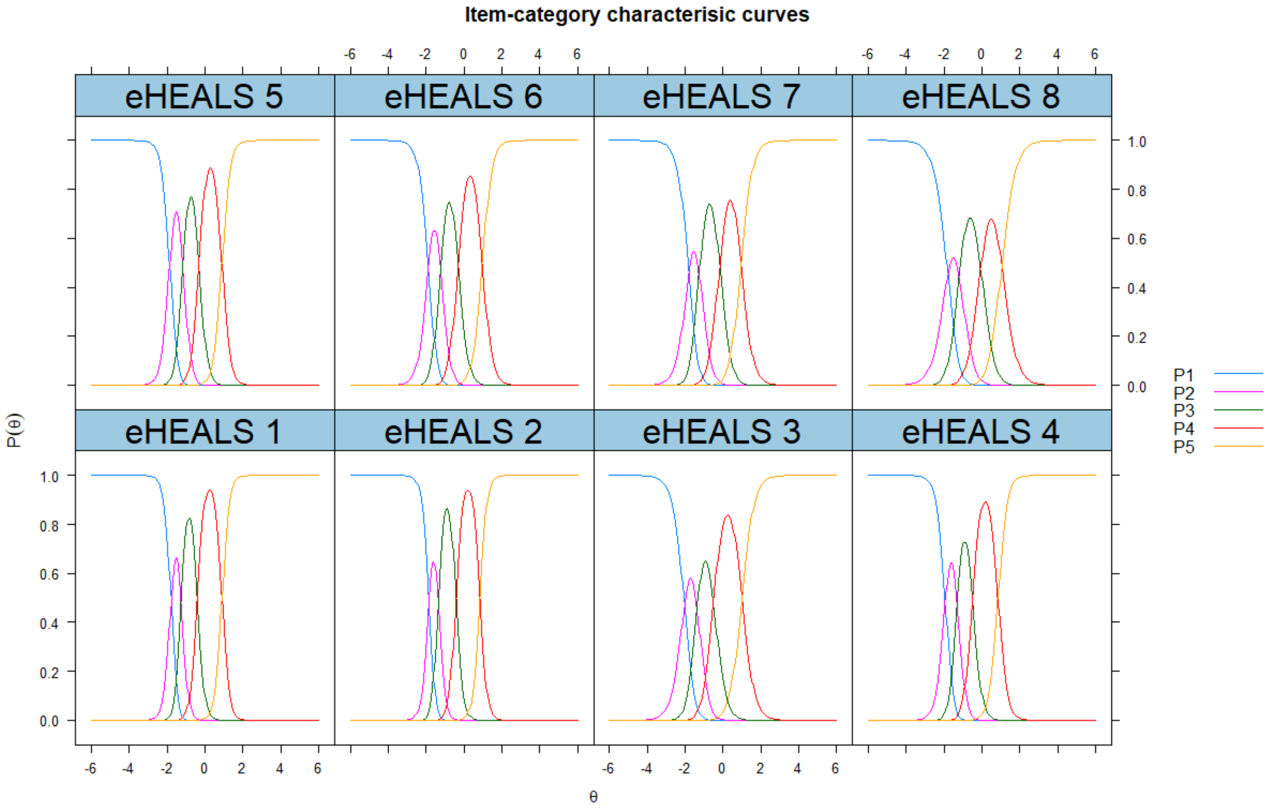


Figure 2. The item information curves for items of the simplified Chinese eHealth literacy scale. eHEALS: eHealth literacy scale.

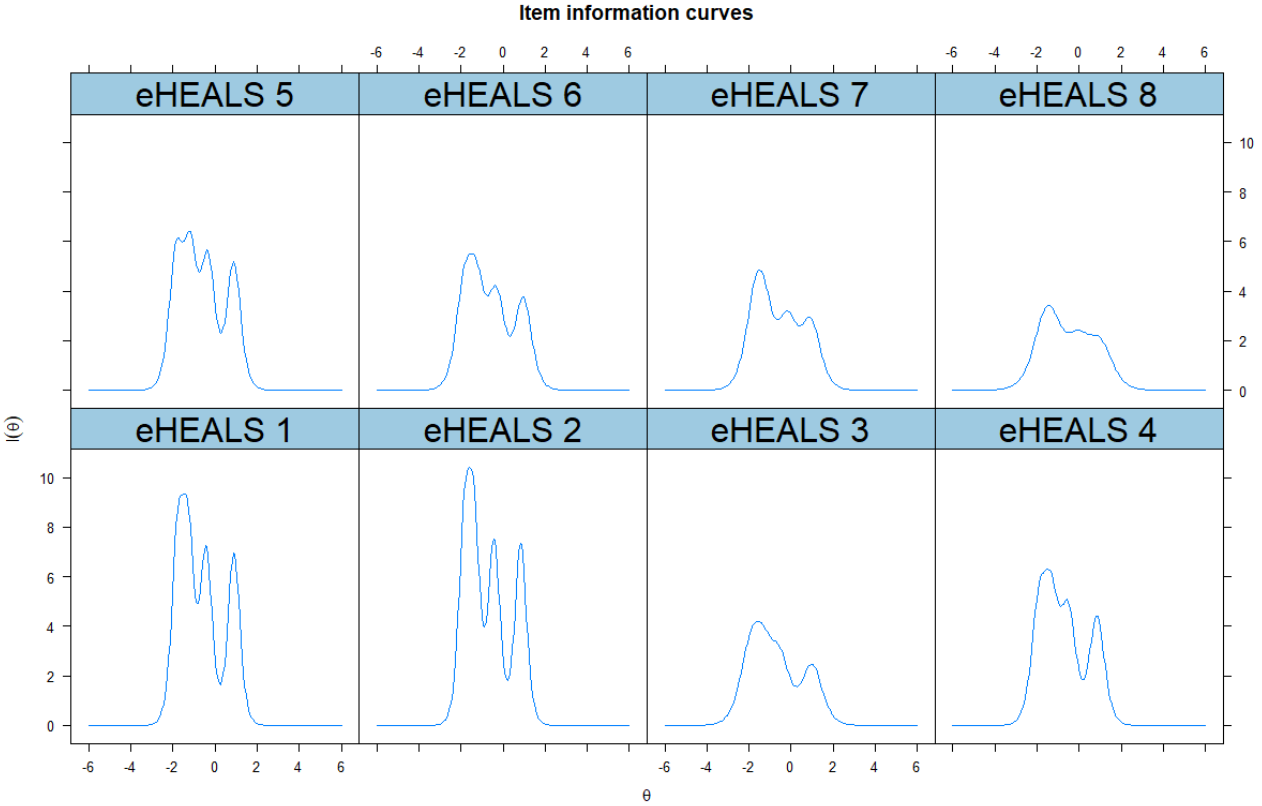
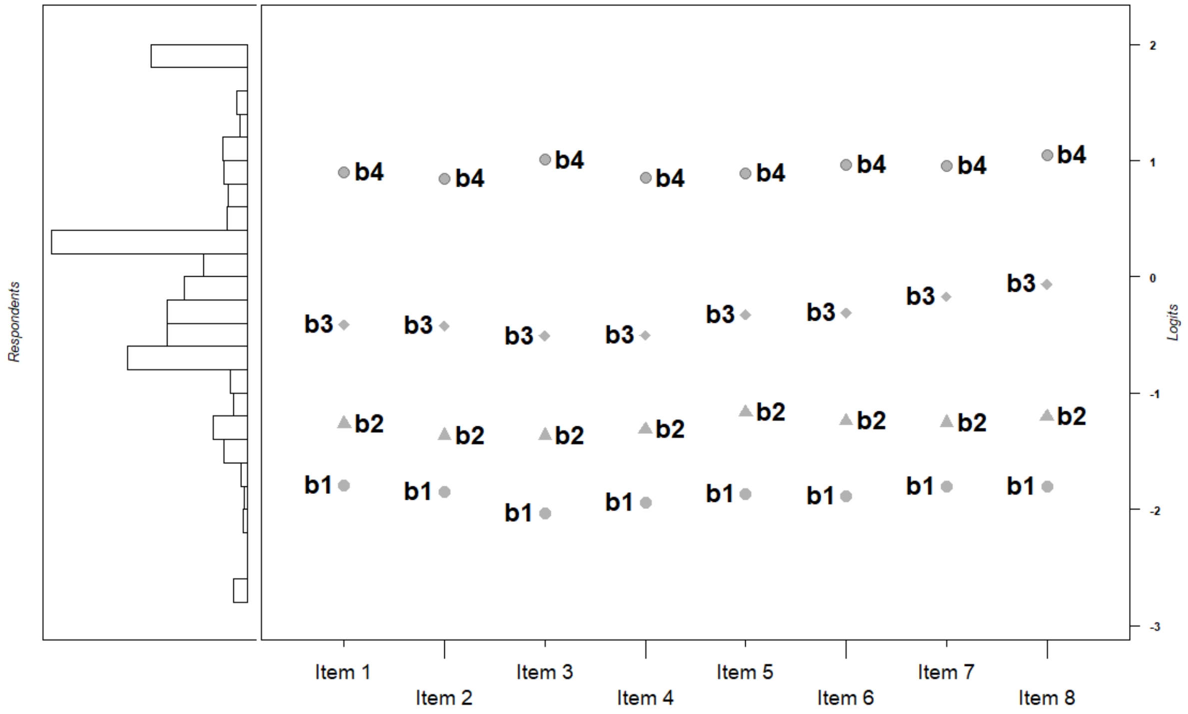


Figure 3. Item-person map of the simplified Chinese eHealth literacy scale. b: coefficient of threshold.



DIF Analysis

Item 6 of the SC-eHEALS showed a uniform DIF when considering the respondents from different age groups. However, the magnitude was smaller than 0.13, confirming that the effect size of the DIF was negligible [29] (Multimedia Appendix 7).

Discussion

Principal Findings

In China, an increasing number of people have turned to the internet to seek health care information because of the rapid proliferation of web-based medical information. However, whether users can leverage such information to improve their health can hardly be measured. The findings of this study support that the SC-eHEALS (Multimedia Appendix 8) is a valid and reliable instrument for measuring self-reported eHealth literacy in China. For the general public, the SC-eHEALS provides a measure to help them evaluate web-based health information and critically appraise the quality of eHealth resources, which could protect them from the harm of low-quality information and empower them to manage their health. Policy makers provide important information for understanding the population's perceived degree of eHealth literacy and developing cost-effective strategies for upgrading the medical care sector by leveraging the internet [30].

Comparisons With Previous Studies

The original one-factor structure of the eHEALS, as confirmed by previous studies in mainland China [13,14], was acceptable, but was not fully supported by our study. This was not an unforeseen result; some other studies have reported a two-factor [4,31] or a three-factor structure [3,32]. Diviani et al [33] pinpointed that when using CFA, neither the single-factor model, originally proposed by the developers, nor the two-factor model, suggested by other research, showed an adequate fit to the data. Cummings [34] indicated that translation might change the original meaning of the items, and these changes could affect perceived meanings for targeted respondents. Furthermore, the original eHEALS was developed at a time before the rise of social media, which totally changed people's interaction with health information, which might affect the structure of the eHEALS [32]. Furthermore, the eHEALS was developed based on a model that originally consisted of 6 domains of literacy, and Noman et al [12] suggested that each skill would require independent measurement. However, in this study, the two-factor structure was not without problems. First, the values of the CFI, TLI, and SRMR indicated that the performance of the revised one-factor model was not worse than that of the two-factor model. Only the RMSEA value supported the revised two-factor model outperforming the other models. A similar finding was reported in the study by Paige et al [3], in which they finally confirmed the one-factor structure. Second, the factor loadings of each item in the revised two-factor model were similar to the one-factor model; item 7 even showed a lower value (0.594), which indicated that the difference in stability between the 2 structures was negligible. Given that people's different abilities to manipulate web-based information is attributed to their demographics, SES, and health conditions, we decided to maintain the one-factor structure. Studies exploring the structure of the SC-eHEALS in other subpopulations should be carried out in the future. A clear structure of the SC-eHEALS would be useful in facilitating the computer adaptive test (CAT) in future practice. Administering the SC-eHEALS based on CAT can strengthen its precision and sensitivity [3], which ensures that the chosen items can meet the respondents' eHealth literacy levels.

The proportion of respondents choosing the options of *neutral* and above was high in this study, which might indicate that most of them are equipped with middle-to-high eHealth literacy and skills and are confident in searching, understanding, analyzing, and using eHealth information. The ceiling effect of all the items of the SC-eHEALS was detected, despite the strength of the effect not being very strong. The distribution of the responses in eHEALS was not reported by other studies in China. However, Paige et al [3] reported that the mean score of the eHEALS ranged between 3.57 and 3.96 in the US population. It remained that the discriminant ability of the SC-eHEALS might not be strong enough to differentiate people with different levels of eHealth literacy. In addition, the SC-eHEALS showed an excellent reliability with a Cronbach alpha (.96) higher than that reported by other studies conducted in China [13,14] and for some other language versions [3,4,12,35,36]. However, Chang and Schultz [13] found that removing items 7 or 8 could improve the reliability of the eHEALS in China, which was not reported by any other study. Moreover, we found that the mean score of the eHEALS varied widely across different populations. For example, Diviani et al [33] reported the same findings as ours, as respondents rated item 4 as the easiest and item 8 as the most difficult. Chung and Nahm [36] found that item 4 was perceived as the easiest item, whereas item 7 was perceived as the most difficult item by a US sample. Van der Vaart et al [4] identified that items 3 and 8 received the highest and lowest mean scores of all items, respectively. Given that different studies have recruited samples with different subpopulations and that the proliferation of the internet in different countries is varied, further exploration should be carried out to draw more international comparisons.

The GPCM was demonstrated to be the best-fit IRT model in this study. No previous studies used the GPCM to test the psychometric properties of eHEALS. Diviani et al [33] used the nonparameter IRT model, Paige et al [3] used the PCM model, and Zrubka et al [35] and Ma and Wu [14] used the graded response model. Our study enriches the knowledge of the application of IRT models and supports existing research on the psychometric analysis of eHEALS using IRT methods. GPCM analysis showed that the overall performance of SC-eHEALS was satisfactory. ICCs indicated that the response categories of each item were ordered, and all categories were most probably at the same point on the continuum. For the location parameter, all items were placed between −3 and 3 across the scale of the latent trait. The discrimination of the items was positive and could discriminate between individuals with different levels of eHealth literacy. The item-person map demonstrated that the items of the SC-eHEALS were located within a reasonable range and were well centered with respect to the person measure distribution. Diviani et al [33] reported similar findings in the Italian eHEALS; however, Paige et al [3] showed that the response categories of the eHEALS covered a wider range of latent traits. The information curves showed a multimodal distribution of rural Chinese citizens [14] and the Hungarian population as shown in the study by Zrubka et al [35]. The TIC of the SC-eHEALS provided the precise estimation near the center of the ability scales, which ranged between −3 and 2, but as the ability level approached the extremes of the scale, the accuracy of the test decreased significantly. The plot of misfit items showed that the majority of imprecise estimations occurred at the options of *neutral* and agree, indicating that these 2 options might not be accurate enough to measure the latent trait among people with moderate-to-high eHealth literacy. To yield a more precise measure of eHealth literacy, future studies should explore what response options are more appropriate to be included in the SC-eHEALS [4].

Moreover, item 6 (skills to evaluate health resources) was labeled as having DIF among respondents of different age groups (older vs younger), which reflected that they have an unequal probability of giving a response to this item. No previous study reported a similar finding at the item level using DIF analysis. Ma and Wu [14] indicated that item 6 has the lowest discriminant power in Chinese eHEALS. However, caution should be exercised when drawing this conclusion, as we cannot determine whether the DIF occurred due to a form of measurement error [37]. In other words, we can neither unequivocally infer that the DIF of item 6 reflected real group differences of the underlying trait among respondents nor attribute this finding to bias during the process of measurement.

Strengths and Limitations

This study has several strengths. First, the respondents who participated in this study came from both developed and underdeveloped areas, were urban and rural residents, and spanned a wide age range, showing better representativeness than the previous studies in China. Second, this study assessed the psychometric properties of SC-eHEALS using both CTT and IRT methods. Third, for IRT analysis, the best-fit model was selected based on comparison of the performance of 3 IRT models. Fourth, DIF analysis was first used to evaluate the item bias and variance of the eHEALS in the Chinese population. The comparisons between the characteristics of the Chinese eHEALS validation studies are presented in Multimedia Appendix 9.

Several limitations of this study should be addressed. First, all respondents were from inpatient departments. This means that most of them might have poor health status, which posed some degree of selection bias. Patients with mild or no health problems should be included in follow-up studies to further test the psychometric properties of SC-eHEALS. Second, the sample size of this study was less than 1000, which might generate some uncertainties in estimating the IRT model, especially using the GPCM (a two-parameter model). A larger sample should be collected in future studies to validate our findings. Third, we did not differentiate between patients with different diseases when evaluating the psychometric properties of the SC-eHEALS, which might have created some problems in explaining the validity of the instrument. The performance of SC-eHEALS should be further assessed in distinct patient groups. Fourth, we did not collect information on respondents' behaviors related to their internet use, such as the frequency and types of websites; thus, criterion validity cannot be assessed.

Finally, the information of patients who refused to participate in the survey was not recorded, which might have led to a measure of information bias.

Conclusions

This study evaluated the psychometric properties of SC-eHEALS using a large sample of patients from a multicenter study in China. It provides empirical evidence that SC-eHEALS is a valid, reliable, and parsimonious instrument for evaluating the degree of eHealth literacy in Chinese people with different demographics, SES (eg, rural and urban residents), and health status. CFA did not fully support the original one-factor structure, and further exploration is needed. IRT analysis suggested that SC-eHEALS might not be effective for use in people with very high or low eHealth literacy. Further studies are needed to detect the heterogeneity of the SC-eHEALS in different subpopulations and further assess its criterion validity.

Authors' Contributions

All authors contributed to the concept and study design. RX and SL prepared the translation, RX conducted the data analysis, and RX and LZ drafted the manuscript. All authors critically revised the manuscript. WD and EW supervised the study. All authors approved the final manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The results of the exploratory factor analysis.
[DOCX File , 20 KB - jmir_v22i12e18613_app1.docx]

Multimedia Appendix 2
The standardized factor loadings for the observed variables of all confirmatory factor analysis models.
[DOCX File , 23 KB - jmir_v22i12e18613_app2.docx]

Multimedia Appendix 3
The simplified Chinese eHealth literacy scale scores stratified according to respondents’ characteristics.
[DOCX File , 21 KB - jmir_v22i12e18613_app3.docx]

Multimedia Appendix 4
The comparisons between different item response theory models.
[DOCX File , 20 KB - jmir_v22i12e18613_app4.docx]

Multimedia Appendix 5
The misfit items of the simplified Chinese eHealth literacy scale.
[PNG File , 682 KB - jmir_v22i12e18613_app5.png]

Multimedia Appendix 6
Test information curve of the simplified Chinese eHealth literacy scale.
[PNG File , 131 KB - jmir_v22i12e18613_app6.png]

Multimedia Appendix 7
Results of differential item functioning analysis.
[DOCX File , 20 KB - jmir_v22i12e18613_app7.docx]

Multimedia Appendix 8
The Chinese version of the eHealth literacy scale.
[DOCX File , 22 KB - jmir_v22i12e18613_app8.docx]

Multimedia Appendix 9
Summary of the Chinese eHEALS studies.
[DOCX File , 21 KB - jmir_v22i12e18613_app9.docx]

References

1. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
2. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. Comput Hum Behav 2017 Apr;69:256-267. [doi: 10.1016/j.chb.2016.12.032]
3. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: an item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb;100(2):320-326 [FREE Full text] [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
4. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 9;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
5. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan 13;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
6. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the ehealth literacy assessment toolkit. J Med Internet Res 2018 May 10;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
7. Parker S, Prince A, Thomas L, Song H, Milosevic D, Harris MF. Electronic, mobile and telehealth tools for vulnerable patients with chronic disease: a systematic review and realist synthesis. BMJ Open 2018 Aug 29;8(8):e019192. [doi: 10.1136/bmjopen-2017-019192] [Medline: 30158214]
8. Neter E, Brainin E. Perceived and performed ehealth literacy: survey and simulated performance test. JMIR Hum Factors 2017 Jan 17;4(1):e2. [doi: 10.2196/humanfactors.6523] [Medline: 28096068]
9. Norman CD, Skinner HA. Ehealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
10. China's H2 2018 Internet User Base Reaches 829 million. China Telecom. 2019. URL: https://www.marbridgeconsulting.com/marbridgedaily/archive/article/114539/chinas_h2_2018_internet_user_base_reaches_829_mln [accessed 2019-02-01]
11. Development and Promotion of 'Internet + Healthcare' in China. State Council of China. 2018. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2019-02-01]
12. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
13. Chang A, Schulz P. The measurements and an elaborated understanding of Chinese ehealth literacy (C-EHEALS) in chronic patients in China. Int J Environ Res Public Health 2018 Jul 23;15(7):- [FREE Full text] [doi: 10.3390/ijerph15071553] [Medline: 30041420]
14. Ma Z, Wu M. The Psychometric Properties of the Chinese eHealth Literacy Scale (C-eHEALS) in a Chinese Rural Population: Cross-Sectional Validation Study. J Med Internet Res 2019 Oct 22;21(10):e15720 [FREE Full text] [doi: 10.2196/15720] [Medline: 31642811]
15. Zibrik L, Khan S, Bangar N, Stacy E, Novak Lauscher H, Ho K. Patient and community centered eHealth: exploring eHealth barriers and facilitators for chronic disease self-management within British Columbia's immigrant Chinese and Punjabi seniors. Health Policy and Technology 2015 Dec;4(4):348-356. [doi: 10.1016/j.hlpt.2015.08.002]
16. Guo S, Yu X, Sun Y, Nie D, Li X, Wang L. eHEALS. Zhongguo jian kang jiao yu 2013;29:106-108.
17. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, ISPOR Task Force for TranslationCultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
18. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. Psychol Assess 1995;7(3):286-299. [doi: 10.1037/1040-3590.7.3.286]
19. Klein B. Principles and Practice of Structural Equation Modeling. Fourth Edition. New York, USA: The Guilford Press; 2015.
20. Linacre J. Sample Size and Item Calibration or Person Measure Stability. Rasch Meas Trans. 1994. URL: http://www.rasch.org/rmt/rmt74m.htm [accessed 2019-02-01]
21. Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Qual Life Res 2012 Mar;21(1):167-176 [FREE Full text] [doi: 10.1007/s11136-011-9927-2] [Medline: 21598064]
22. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Eq Model: Multidiscip J 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
23. Xie Y, DeVellis RF. Scale development: theory and applications. Contemp Sociol 1992 Nov;21(6):876. [doi: 10.2307/2075704]
24. DeMars C. Item Response Theory. Oxford, UK: Oxford University Press; 2010.

25. Coulacoglou C. Psychometrics and Psychological Assessment: Principles and Applications. London, UK: Academic Press; 2017.
26. Valdez JP, Chu SK. Examining the psychometric validity of the five-item gratitude questionnaire: an item response theory approach. J Psychoeduc Assess 2018 Dec 17;38(4):529-536. [doi: 10.1177/0734282918816542]
27. Xu RH, Cheung AW, Wong EL. Development and validation of an instrument to measure patient engagement in Hong Kong Special Administrative Region, China. Patient Prefer Adherence 2018 Sep;Volume 12:1667-1675. [doi: 10.2147/ppa.s171026]
28. Zumbo A. A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binaryand Likert-type (Ordinal) Item Scores. Semantic Scholar. 1999. URL: https://www.semanticscholar.org/paper/A-Handbook-on-the-Theory-and-Methods-of-Item-(DIF)-Zumbo/7f88fb0ad98645582665532600d7c46406fa2db6 [accessed 2019-02-01]
29. Lipovetsky S. Modern psychometrics With R. Technometrics 2020 Jan 31;62(1):135-137. [doi: 10.1080/00401706.2019.1708675]
30. Neter E, Brainin E, Baron-Epel O, editors. The third digital divide in the health domain: is internet use for health purposes associated with health benefits? In: EHealth: Current Evidence, Promises, Perils, and Future Directions. Bingley, UK: Emerald Publishing Limited; 2018.
31. Soellner R, Huber S, Reder M. The concept of ehealth literacy and its measurement. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
32. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the measurement properties of the ehealth literacy scale (EHEALS) among baby boomers: a multinational test of measurement invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
33. Diviani N, Dima AL, Schulz PJ. A psychometric analysis of the Italian version of the ehealth literacy scale using item response and classical test theory methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
34. Cross-cultural survey methods. Technometrics 2003 May;45(2):182-182. [doi: 10.1198/tech.2003.s151]
35. Zrubka Z, Hajdu O, Rencz F, Baji P, Gulácsi L, Péntek M. Psychometric properties of the Hungarian version of the eHealth Literacy Scale. Eur J Health Econ 2019 Jun;20(Suppl 1):57-69 [FREE Full text] [doi: 10.1007/s10198-019-01062-1] [Medline: 31098883]
36. Chung S, Nahm E. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156 [FREE Full text] [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
37. Santelices MV, Wilson M. On the relationship between differential item functioning and item difficulty. Educ Psychol Measur 2011 Aug 4;72(1):5-36. [doi: 10.1177/0013164411412943]

Abbreviations

AIC: Akaike information criterion
ANOVA: one-way analysis of variance
BIC: Bayesian information criterion
CAT: computer adaptive test
CFA: confirmatory factor analysis
CFI: comparative fit index
CTT: classical test theory
DIF: differential item functioning
EFA: exploratory factor analysis
eHEALS: eHealth literacy scale
GPCM: general partial credit model
ICC: item-category characteristic curve
ICECAP-A: ICEpop capability measure for adults
IIC: item information curve
IRT: item response theory
PCC: patient-centered care
PCM: partial credit model
RMSEA: root mean square error of approximation
SC-eHEALS: simplified Chinese eHealth literacy scale
SES: socioeconomic status
SRMR: standardized root mean square residual
TIC: test information curve
TLI: Tucker-Lewis index
VAS: visual analogue scale

Edited by G Eysenbach, R Kukafka; submitted 11.03.20; peer-reviewed by S Mitsutake, E Neter, Y Chu, M Fiordelli; comments to author 09.09.20; revised version received 25.09.20; accepted 15.11.20; published 07.12.20.

Please cite as:
Xu RH, Zhou L, Lu SY, Wong EL, Chang J, Wang D
Psychometric Validation and Cultural Adaptation of the Simplified Chinese eHealth Literacy Scale: Cross-Sectional Study
J Med Internet Res 2020;22(12):e18613
URL: http://www.jmir.org/2020/12/e18613/
doi:10.2196/18613
PMID:33284123

Original Paper

Relationship Between Levels of Digital Health Literacy Based on the Taiwan Digital Health Literacy Assessment and Accurate Assessment of Online Health Information: Cross-Sectional Questionnaire Study

Peggy Liu[1,2], MSMT; Ling-Ling Yeh[3], PhD; Jiun-Yi Wang[1,4], PhD; Shao-Ti Lee[2], BSc

[1]Department of Healthcare Administration, Asia University, Taichung, Taiwan

[2]International Federation for Information Integration, Taipei, Taiwan

[3]Social Enterprise and Innovation MA Program, Graduate School of Humanities and Social Sciences, Dharma Drum Institute of Liberal Arts, New Taipei City, Taiwan

[4]Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan

Corresponding Author:
Ling-Ling Yeh, PhD
Social Enterprise and Innovation MA Program
Graduate School of Humanities and Social Sciences
Dharma Drum Institute of Liberal Arts
No 700, Fagu Rd, Jinshan Dist
New Taipei City, 20842
Taiwan
Phone: 886 224980707 ext 5205
Email: YehLL@dila.edu.tw

Abstract

Background: The increasing amount of health information available on the internet makes it more important than ever to ensure that people can judge the accuracy of this information to prevent them from harm. It may be possible for platforms to set up protective mechanisms depending on the level of digital health literacy and thereby to decrease the possibility of harm by the misuse of health information.

Objective: This study aimed to create an instrument for digital health literacy assessment (DHLA) based on the eHealth Literacy Scale (eHEALS) to categorize participants by level of risk of misinterpreting health information into high-, medium-, and low-risk groups.

Methods: This study developed a DHLA and constructed an online health information bank with correct and incorrect answers. Receiver operating characteristic curve analysis was used to detect the cutoff value of DHLA, using 5 items randomly selected from the online health information bank, to classify users as being at low, medium, or high risk of misjudging health information. This provided information about the relationship between risk group for digital health literacy and accurate judgement of online health information. The study participants were Taiwanese residents aged 20 years and older. Snowball sampling was used, and internet questionnaires were anonymously completed by the participants. The reliability and validity of DHLA were examined. Logistic regression was used to analyze factors associated with risk groups from the DHLA.

Results: This study collected 1588 valid questionnaires. The online health information bank included 310 items of health information, which were classified as easy (147 items), moderate (122 items), or difficult (41 items) based on the difficulty of judging their accuracy. The internal consistency of DHLA was satisfactory (α=.87), and factor analysis of construct validity found three factors, accounting for 76.6% of the variance. The receiver operating characteristic curve analysis found 106 people at high risk, 1368 at medium risk, and 114 at low risk of misinterpreting health information. Of the original grouped cases, 89.6% were correctly classified after discriminate analysis. Logistic regression analysis showed that participants with a high risk of misjudging health information had a lower education level, lower income, and poorer health. They also rarely or never browsed the internet. These differences were statistically significant.

Conclusions: The DHLA score could distinguish those at low, medium, and high risk of misjudging health information on the internet. Health information platforms on the internet could consider incorporating DHLA to set up a mechanism to protect users from misusing health information and avoid harming their health.

(J Med Internet Res 2020;22(12):e19767) doi:10.2196/19767

KEYWORDS

digital health literacy; internet health information; risk group; Taiwan

Introduction

The International Telecommunication Union reported in 2019 that over one-half of the world's population was using the internet. In total, 97% of the world's population lives within reach of a mobile cellular signal and 93% lives within reach of a network that provides a 3G signal or better [1]. In 2014, a European Union report stated that approximately 60% of the public browses the internet every day and 60% of those use the internet to search for health information [2]. More than 70% of US adults used the internet as their primary source of health information in 2019 [3]. In Taiwan, the 2016 Household Digital Opportunity Survey Report [4] concluded that 66.5% of internet users searched for health education, health, or food safety information on the internet [4]. A multinational study by Song et al [5] reported that 51.5% of Americans, 76.9% of South Koreans, and 81.4% of Hongkongers reported using social networking sites to obtain health information, and 66.2%, 94.6%, and 86.1% of them, respectively, reported using content from health information blogs. This implies that people may use the internet to transmit or retrieve health information at any time and anywhere, with the rapid development of digital information tools.

However, Lee et al [6] reported that technology and medical internet rumors accounted for 30% of the total rumor volume. In 2019, a survey in Taiwan found that 26.8% of internet users aged 12 years and older received and forwarded unverified information or news, and 44.9% did not verify information or news [7]. People also tend to believe health information forwarded by friends and relatives because they trust the forwarder [8]. Unnecessary harm could be avoided if people’s digital health literacy could be assessed as a way to understand their capacity to make accurate judgements about online health information. It is hoped that website platforms might then provide warnings about health information that have not been fact-checked or simply do not display that information for people with lower levels of digital health literacy, who are likely to be at high risk of misjudging the accuracy of health information.

The eHealth Literacy Scale (eHEALS) is a subjective measurement that measures perceived skills at finding, evaluating, and employing electronic health information to solve health problems [9,10]. The eHEALS is a widely used measurement tool that has been translated and validated in different languages [11]. Previous studies have used the eHEALS to assess the ability of digital health literacy related to online health information–seeking. Ghaddar et al [12] surveyed online health information–seeking behavior using a questionnaire about patterns of internet use, checking health information on the internet, and knowledge of a particular health information website (Medline Plus [13]). Two studies [14,15] used telephone interviews to ask participants whether they had ever used the internet, their frequency of use, and where they accessed the internet. Diviani et al [16] carried out qualitative in-depth interviews and explored experiences of online health information–seeking and judgment of credibility of online health information. Noblin and Rutherford [17] asked whether participants used a personal health record or other internet-based information system for health care, and whether they were willing to use health care information technology if they did not already do so. These studies, therefore, explored various aspects of online health information–seeking. However, all of the studies used self-reporting, which is regarded as a subjective measurement.

van der Vaart et al [18] explored the association between eHEALS results and an actual performance test lasting up to 1.5 hours. The performance test incorporated operational, formal, information, and strategic internet skills using objective measurements. For example, how to access a health website, save a file, or add a website to the Favorites menu were all operational internet skills [18]. A study by Quinn et al [19] investigated the relationship between health literacy, eHealth literacy, and actual online health information–seeking behavior by integrating software into a web browser to objectively monitor online interaction; the data were coded for analysis. The study participants were recruited from a convenience sample of 54 university students and staff, and they completed the search tasks under laboratory conditions with 6 health questions about diabetes, obesity, influenza, nutrition, and analgesic medication. These studies, therefore, used objective measures, but their designs were not directly comparable with real-world situations because of participants’ characteristics or the limited health questions or time required. To the best of our knowledge, no study has yet reported on the relationship between the risk of misinterpreting health information, measured as level of digital health literacy, and ability to make accurate judgments about online health information, especially using a wide range of health information and a large number of participants with real-world scenarios.

Therefore, this study aimed to develop an instrument for digital health literacy assessment (DHLA) based on the eHEALS. It distinguished the different risk levels for digital health literacy using DHLA scores by assessing the accuracy of judgments about internet health information resembling that found in the real world.

Methods

We developed an instrument called the DHLA, which drew from the eHEALS. An online health information bank with right and wrong answers was also constructed to check the ability of the DHLA to discriminate between people at low, medium, and high risk of misinterpreting health information using receiver operating characteristic curve (ROC) analysis. This study was reviewed and approved by the regional review board in central Taiwan (approval number CRREC-108-096).

Study Participants

This study's participants were Taiwanese residents aged 20 years and older. Snowball sampling was used to identify potential participants among colleagues, friends, and family members of the researchers in northern, central, and southern Taiwan. A total of 350 participants were recruited, to maximize the range of socioeconomic status. They were asked to send the online survey to their friends and family members. Email and online communications software were used to send quick response codes or website links to participants to enable them to complete the online questionnaire.

Instruments

The eHEALS includes 8 questions that use a 5-point Likert-type scale, with scores ranging from 1 (strongly disagree) to 5 (strongly agree). This generates a single factor. The total score ranges from 8 to 40 points. A higher score shows better digital health literacy [9,10]. A previous study [20] showed that the eHEALS structure varies with the environment and setting, and researchers have also recommended that the scale should consider environmental and cultural issues [21,22]. In Taiwan, Wu [23] used Yahoo data to establish that almost 15% of items of health information obtained during searches related to folklore and custom. An example item was that "smokers should eat pig's blood to cleanse their lungs." Therefore, 2 questions were added in this study. Question 1 asked participants to rate their ability to use information technology ("I am able to use a computer/smartphone to find information that I need on the internet"), and question 10 asked about their belief in local folk medicine ("I feel confident about the health care information based on folklore and customs that I found on the internet"). The responses to the questions directly rated the level of ability and belief in the question content, instead of rating attitudes with levels of agreement. Questions 1 to 6 involved self-assessment of digital health literacy with responses from 1 (very poor) to 5 (very good). Questions 7 to 9 were concerned with how convincing people found internet health information from different sources, and question 10 was about trust in health information from folklore and customs. The responses ranged from 1 (not at all convincing) to 5 (very convincing). The 10 items of the DHLA are shown in Multimedia Appendix 1.

Construction of the Online Health Information Bank

This study constructed a scenario in which internet users could complete an assessment of their digital health literacy using the DHLA before starting their search for health information. This would allow the website to identify their level of digital health literacy. Users with a lower level of digital health literacy would then see a warning message alongside unverified items of health information, or simply not see that kind of health information. Thus, this study involved a search of items of health information from the internet in the real world. The health sectors in Taiwan set up a platform to conduct fact-checking of health information issued by people on the internet. Items of health information were obtained from the Fact Check Column on the Taiwan Food and Drug Administration website [24] and the Healthcare Fact Check Column on the Health Promotion Administration website [25]. These sites carry information about food safety, drugs, medical devices, cosmetics, medical treatment, preventative health, disease screening, health promotion, and general medicine. Any resident of Taiwan can check this health information via an internet search. The health authority invites experts to meet and develop a consensus about whether the information is right or wrong, with a rationale. The content of items of health information and expert feedback were obtained from public government information platforms [26,27]. The online health information bank in this study focused on food safety, drugs, medical treatment, preventative health, disease screening, health promotion, and general medicine, which is very similar to the range of health information available on the internet.

We carried out a two-stage content validity check to construct the online health information bank. A total of 600 items of health information were included from the two public sector platforms, which was reduced to 529 after repeated items were removed. In stage 1, four medical experts were invited to classify how difficult it was to judge the accuracy of each piece of content and feedback from the public platforms. The items were classified into categories as easy, moderate, or difficult to judge, defined as items for which more than 80%, 50%, and less than 30% of the public, respectively, were able to determine credibility. In stage 2, one public health expert and three members of the public were invited to discuss the expert difficulty categorizations. The group meeting considered that the items of health information must be stated precisely, and that highly specialized items of health information were not appropriate. Therefore, items of health information were excluded if experts had judged them to be only partially correct, or where determining credibility required knowledge of chemical composition and its effects. This resulted in an online health information bank containing 310 items, of which 147 were easy, 122 were moderate, and 41 were difficult. Sample items and difficulty levels are included in Multimedia Appendix 2.

Data Collection

The questionnaire was completed anonymously via an online questionnaire platform. The questionnaire introduced the DHLA, then asked for basic information such as sex, age group, education level, marital status (married or not), employment, income, whether their job was health-related, quantity (level) of internet use, self-reported health, height, weight, waist circumference, exercise frequency, and place of residence. Finally, the system randomly selected 5 items (2 easy, 2 moderate, and 1 difficult) from the online health information bank for participants to consider. For each item, participants were asked to determine whether the information was right or wrong, with an option for unsure. The questionnaire platform set standard answers, and participants were given a score of 1 when the answer given was correct (accurate judgment), and 0 when they answered incorrectly or were not sure (not accurate judgment). Items were selected at random because real-world health information is varied and has diverse levels of difficulty. This approach therefore simulated a real-world search for health information on the internet. Once the participant had completed the questionnaire, the platform showed them the correct answers to the 5 items for their future reference, as an important ethical issue.

Analysis

The reliability and internal consistency of the DHLA was assessed using Cronbach alpha and item-scale correlation. The validity was assessed using convergent validity and construct validity. Spearman correlations were used between total DHLA scores and age, education level, and level of internet use. In line with previous studies, we hypothesized negative correlations with age and positive correlations with education level and level of internet use, and these were used to investigate the convergent validity [16,18]. Principal component analysis with promax rotation was used for construct validity, assuming that there were mild correlations among the factors, and eigenvalues greater than 1 were used as an indication of how many factors to retain. Distributional properties of the DHLA were investigated by examining the normality of the total scores and detecting floor and ceiling effects. Skewness and kurtosis values below −1 or above 1 were considered to show non-normality [28]. Floor or ceiling effects were considered present if more than 15% of participants scored either the worst or the best possible score on the DHLA [29].

We used ROC analysis to distinguish high-, medium-, and low-risk groups using different cutoff points for the DHLA score, based on the participants’ ability to make an accurate judgment of easy, moderate, and difficult items. A cutoff value for DHLA scores was determined from the point on the ROC curve. In general, the cutoff point is based on an area of more than 50% under the curve. We used the Youden index to select the most appropriate cutoff value: (sensibility + specificity – 1) is maximum [30].

Descriptive statistics were used to demonstrate the proportion of correct answers, with means and standard deviations of the DHLA scores in different risk groups. Chi-square tests were used to detect the relationship between difficulty levels of the health information items and the risk groups. Analysis of variance (ANOVA) was used to investigate the differences in the DHLA score for each factor across different risk groups. Posthoc analysis was used to test whether the differences among risk groups were statistically significant following ANOVA. We used discriminate analysis on the three risk groups to explore the probability of correctly classifying people by their DHLA scores. Convergent validity was tested using Spearman correlation analysis to examine the relationship between difficulty levels of items of health information and risk groups, in terms of age, education level, and level of internet use.

In the last stage of data analysis, the chi-square test was used to analyze the relationship between digital health literacy and individual attributes, self-reported health status, and level of internet use, followed by logistic regression analysis to analyze factors related to risk group identified by DHLA scores. The high-risk group was compared with the medium-risk group and with a combined medium- and low-risk group. SPSS statistical software (version 22.0; IBM Corp) was used for statistical analysis. *P* values <.05 were considered statistically significant.

Results

Participants

A total of 1871 online questionnaires were collected, of which 1588 were valid after excluding those that were not complete. In total, 63.7% (1011/1588) of respondents were women, 40.6% (645/1588) were aged 45 to 63 years, 87.7% (1393/1588) had an education level of college (959/1588, 60.4%) or above (434/1588, 27.3%), and 57.8% (918/1588) were married. A total of 46.7% (742/1588) of the participants were professionals, and 36.9% (586/1588) worked in the service industry; 36.1% (573/1588) had a monthly income of New Taiwan dollar (NT) $30,000 to $50,000 (US $1049 to US $1748.34), 91.1% (1447/1588) browsed the internet several times every day, and 46.3% (736/1588) perceived that their health was good. Finally, 15.6% (247/1588) exercised every day and 20.9% (332/1588) did not exercise at all.

Distributional Properties

The total DHLA scores were approximately normally distributed, with a skewness of −0.53 and a mean score of 35.50 (SD 5.56). Nobody obtained the worst possible score (ie, 10) and 5 (5/1588, 0.3%) participants scored the maximum possible score (ie, 50), so the floor and ceiling effects were acceptable.

Reliability and Validity

We used principal component factor analysis on DHLA scores to give three factors. These three factors were digital health literacy (questions 1 to 6), belief in medicine (questions 7 to 9), and belief in folk remedies (question 10), and they accounted for 76.6% of the variance, with eigenvalues of 4.87, 1.68, and 1.11, respectively. Table 1 shows the factor loadings for the 10 items, ranging from 0.63 to 0.93. The standardized Cronbach alpha was .87 for the whole scale, .92 for the subscale on digital health literacy, and .77 for the subscale on belief in medicine. Thus, the internal consistency overall was good. All item-total correlations were also statistically significant, ranging from 0.27 to 0.87. If an item was deleted, Cronbach alpha ranged from .84 to .89.

Table 1. Digital health literacy assessment (DHLA) internal reliability and factor structure.

DHLA item		Mean (SD)	Factor loading Domain name[a] 1	 2	 3	Item-scale correlation, $P<.001$	Cronbach alpha if item deleted
Full scale		35.50 (5.56)					.87
1	My ability to use computer/smartphone to find information that I need on the internet	4.04 (0.91)	0.77			0.73	.85
2	My ability to find health- or disease-related information on the internet	3.89 (0.86)	0.91			0.85	.84
3	My ability to find information on internet to understand health problems or diseases	3.82 (0.86)	0.93			0.87	.84
4	My ability to find information on the internet to answer the questions on health care or disease treatment	3.67 (0.88)	0.91			0.85	.84
5	My ability to use information found on the internet to discuss with health care professionals	3.27 (0.99)	0.79			0.75	.85
6	My ability to judge whether the health care information found on the internet is accurate or not	3.45 (0.88)	0.82			0.78	.85
7	Beliefs about the health care information that I find on the internet	3.31 (0.64)		0.63		0.53	.87
8	Beliefs about the health care information provided by physicians that I find on the internet	3.60 (0.66)		0.90		0.56	.86
9	Beliefs about the health care information provided by hospitals that I find on the internet	3.82 (0.67)		0.89		0.48	.87
10	Beliefs about the health care information based on folklore and customs that I find on the internet	3.37 (0.79)			0.90	0.27	.89

[a] Domain names are as follows: 1=digital health literacy, 2=belief in medicine, 3=belief in folk remedies.

Table 2 shows the convergent validity using Spearman correlations between total DHLA scores and the variables of age, education level, and level of internet use. All correlation analyses were statistically significant, and the correlation coefficients showed mild correlation (<0.3) with age (ρ=–0.19, *P*<.001), education (ρ=0.22, *P*<.001), and level of internet use (ρ=0.17, *P*<.001).

Table 2. Spearman correlation analysis of total scores for the DHLA and scores for accurate judgment across different difficulty levels and risk groups.

Characteristic	DHLA[a] (ρ)	DHLA (*P* value)	Difficulty level scores						Total score (ρ)	Total score (*P* value)	Risk group (ρ)	Risk group (*P* value)
			Easy[b]		Moderate[c]		Difficult[d]					
			ρ	*P* value	ρ	*P* value	ρ	*P* value				
Age (≤34 to ≤65 years)	−0.19	<.001	0.02	.322	0.03	.268	0.02	.344	0.04	.168	−0.18	<.001
Education level (junior high school or below to university or above)	0.22	<.001	0.05	.196	0.05	.031	0.10	<.001	0.09	<.001	0.18	<.001
Level of internet use (never to several times per day)	0.18	<.001	0.04	.245	0.07	.007	0.07	.006	0.08	<.001	0.21	<.001

[a] DHLA: digital health literacy assessment.

[b] Pearson correlation between DHLA and easy level=0.04.

[c] Pearson correlation between DHLA and moderate level=0.07, *P*=.009.

[d] Pearson correlation between DHLA and difficult level=0.17, *P*<.001.

Defining Risk Group

ROC analysis was used to analyze the scores for each factor of the DHLA to identify the cutoff points between the tendency to answer correctly and incorrectly. As shown in Table 3, there were good cutoff points at 15.5, 8.5, and 1.5 for digital health literacy, belief in medicine, and belief in folk remedies, respectively. The area under the curve was less than 0.50 for all difficulty levels for the factor belief in folk remedies, so a score of less than the cutoff value (1.5) was defined as the tendency to answer correctly. However, for the other two factors (digital health literacy and belief in medicine), less than the cutoff value (15.5 and 8.5, respectively) was defined as the tendency to answer incorrectly. The new scores for the three DHLA factors were therefore defined as 1 (tends to be incorrect) or 2 (tends to be correct) based on the ROC cutoff points. Participants were considered to be at low risk of misjudgments if they scored 2 for all three factors, at medium risk if they scored 2 for two of the three factors, and at high risk if one or none of the three factor scores was 2. This gave a total number of 106 participants in the high-risk group, 1368 participants in the medium-risk group, and 144 participants in the low-risk group.

Table 3. Receiver operating characteristic curve analysis of the cutoff points of the digital health literacy assessment scores by accurate judgment and difficulty level.

Domain name and difficulty level	Cutoff	Sensitivity	Specificity	AUC[a]	95% CI
Digital health literacy					
Easy	15.5	0.95	0.09	0.49	0.45-0.54
Moderate	15.5	0.95	0.09	0.54	0.50-0.57
Difficult	15.5	0.96	0.08	0.56	0.53-0.59
Belief in medicine					
Easy	8.5	0.96	0.03	0.52	0.47-0.56
Moderate	8.5	0.97	0.05	0.50	0.47-0.54
Difficult	8.5	0.97	0.04	0.53	0.50-0.56
Belief in folk remedies					
Easy	1.5	0.91	0.07	0.48	0.44-0.52
Moderate	1.5	0.91	0.10	0.46	0.43-0.50
Difficult	1.5	0.91	0.08	0.48	0.46-0.53

[a] AUC: area under the curve.

Table 4 shows that 86.9% of easy items, 78.3% of moderate items, and 54.2% of difficult items were judged correctly. The highest proportion of correct answers was always in the low-risk group (90.4%, 79.8%, 57.0%), then the medium-risk group (87.0%, 78.9%, 55.0%), and finally the high-risk group (82.1%, 68.9%, 41.5%). There was no statistically significant relationship between correctly assessing the easy items and risk group (*P*=.190). However, there was a statistically significant relationship with risk group for both the moderate and high difficulty items (*P*=.049; *P*=.023). The mean DHLA score in the low-risk group (mean 39.32, SD 4.72) was greater than in the medium- (mean 36.84, SD 4.37) and high-risk groups (mean 25.31, SD 5.04), and this difference was statistically significant. The three factors of digital health literacy, belief in medicine, and belief in folk remedies showed very similar results. For the pair comparison of the post hoc test, the mean scores for digital health literacy and belief in medicine for the low-risk group were both greater than for the medium-risk group, but this difference was not statistically significant. The discriminant analysis showed that 89.6% of the original grouped participants were correctly classified.

Table 4. Comparison of proportion of accurate judgments of online health information sources and digital health literacy assessment (DHLA) scores across different risk groups.

Difficulty level/DHLA	Risk group						
	High (n=106)	Medium (n=1368)	Low (n=114)	Total (n=1588)	χ^2/F (*df*)	*P* value	Posthoc test
Difficulty level (correct answer, %)							
Easy	82.1	87.0	90.4	86.9	3.37 (2)	.190	
Moderate	68.9	78.9	79.8	78.3	6.05 (2)	.049	
Difficult	41.5	55.0	57.0	54.2	7.57 (2)	.023	
DHLA score							
Mean	25.31	36.84	39.32		360.12 (2)	<.001	H[a]<M[b], H<L[c], M<L; *P*<.001
SD	5.04	4.37	4.72				
Digital health literacy score					303.77 (2)	<.001	
Mean	13.14	22.75	23.29				H<M, H<L; *P*<.001
SD	4.60	3.82	4.13				
Belief in medicine					92.36 (2)	<.001	
Mean	8.78	10.86	11.04				H<M, H<L; *P*<.001
SD	2.38	1.46	1.44				
Belief in folk remedies					388.63 (2)	<.001	
Mean	3.39	3.23	5.0				H<M, *P*=.02; H<L, M<L, *P*<.001
SD	0.82	0.66	0.0				
Predicate group[d]							
High	56	50	0				89.6% of original grouped cases correctly classified
Medium	1	1367	0				
Low	0	114	0				

[a] High-risk group.

[b] Medium-risk group.

[c] Low-risk group.

[d] Standardized canonical discriminant function: Wilks lambda=0.688, *P*<.001; eigenvalue=0.454; canonical correlation=0.559.

This study also examined the convergent validity of scores of difficulty level and risk groups in Table 2. The Spearman correlation coefficients across difficulty levels showed a very mild correlation with education level (ρ=0.05–0.10) and level of internet use (ρ=0.04–0.07). The correlations with moderate and difficult items were statistically significant. The Pearson correlation coefficient between the DHLA score and the score in each difficulty level ranged from 0.4 to 0.1 and was statistically significant for moderate and difficult items. The scores for easy items across risk groups were not statistically significant. The correlation coefficients between risk groups and age, education level, and level of internet use indicated a mild but statistically significant correlation (ρ=−0.18 to 0.21). There was a negative correlation with age, and positive correlations with levels of education and internet use.

Description and Associated Factors Analysis in Different Risk Groups

Table 5 shows that gender, age, education level, marital status, employment, income, level of internet use, self-reported health status, and exercise habits were statistically significantly associated with risk group. Overall, 70.7% (75/106) of the subjects aged 45 years and older were in the high-risk group, and 28.9% (33/114) were in the low-risk group. In total, 55.7%

(59/106), 89.5% (1224/1368), and 96.5% (110/114) of participants in the high-, medium-, and low-risk groups, respectively, had a college degree or higher. This shows that higher levels of education were associated with lower risk of misinterpreting health information. Participants who worked as professionals and in the service industry were mostly in the medium- and low-risk groups (703/742 [94.7%] and 556/586 [94.9%]). Participants who worked as laborers accounted for 16.0% (17/106), 6.9% (95/1368), and 2.6% (3/114) of the high-, medium-, and low-risk groups, respectively. Participants who worked in the service industry accounted for 28.3% (30/106), 37.6% (514/1368), and 36.8% (42/114) of participants in the high-, medium-, and low-risk groups, respectively. The proportion of participants with an income of NT $50,000 (US $1748.34) or greater per month was 28.3% (30/106) in the high-risk group, 37.6% (514/1368) in the medium-risk group, and 39.4% (45/114) in the low-risk group, indicating that a higher income is associated with a lower proportion of participants in the high-risk group. The proportion of participants who never or only rarely browsed the internet accounted for 28.3%, 2.5%, and 0% of the high-, medium-, and low-risk groups. A high proportion of the low-risk group (70/114, 61.4%) had good health, and the proportion with poor health in the high-, medium-, and low-risk groups was 17.0% (18/106), 6.6% (90/1368), and 3.5% (4/114). This suggests that most participants with poor health were in the high-risk group. The proportion of participants who did not exercise was higher in the higher risk groups (27/106 [25.5%], 286/1368 [20.9%], and 19/114 [16.7%] in the high-, medium-, and low-risk groups, respectively).

Table 5. Risk groups and characteristic distribution of study participants.

Characteristic	Risk group, n (%)					
	High (n=106)	Medium (n=1368)	Low (n=114)	Total, n (%)	χ^2 (*df*)	*P* value
Sex					17.5 (2)	<.001
Male	46 (43.4)	471 (34.4)	60 (52.6)	577 (36.3)		
Female	60 (56.6)	897 (65.6)	54 (47.4)	1011 (63.7)		
Age (years)					142.6 (6)	<.001
≤34	20 (18.9)	364 (26.6)	52 (45.6)	436 (27.5)		
35–44	11 (10.4)	398 (29.1)	29 (25.4)	438 (27.6)		
45–64	49 (46.2)	564 (41.2)	32 (28.1)	645 (40.6)		
≥65	26 (24.5)	42 (3.1)	1 (0.9)	69 (4.3)		
Education level					234.5 (6)	<.001
Junior high school or below	21 (19.8)	9 (0.7)	0 (0)	30 (1.9)		
Senior high school	26 (24.5)	135 (9.9)	4 (3.5)	165 (10.4)		
University or college	45 (42.5)	841 (61.5)	73 (64)	959 (60.4)		
Graduate school	14 (13.2)	383 (28)	37 (32.5)	434 (27.3)		
Married					7 (2)	.029
No	40 (37.7)	569 (41.6)	61 (53.5)	670 (42.2)		
Yes	66 (62.3)	799 (58.4)	53 (46.5)	918 (57.8)		
Employment					40.8 (6)	<.001
Professional	39 (36.8)	653 (47.7)	50 (43.9)	742 (46.7)		
Service industry	30 (28.3)	514 (37.6)	42 (36.8)	586 (36.9)		
Manual work	17 (16.0)	95 (6.9)	3 (2.6)	115 (7.2)		
Unemployed	20 (18.9)	106 (7.7)	19 (16.7)	145 (9.1)		
Income (NT $)[a,b]					90.3 (12)	<.001
0	22 (20.8)	86 (6.3)	15 (13.2)	123 (7.7)		
≤\$23,800	20 (18.9)	79 (5.8)	18 (15.8)	117 (7.4)		
\$23,801-\$30,000	10 (9.4)	167 (12.2)	9 (7.9)	186 (11.7)		
\$30,001-\$50,000	24 (22.6)	522 (38.2)	27 (23.7)	573 (36.1)		
\$50,001-\$70,000	12 (11.3)	276 (20.2)	18 (15.8)	306 (19.3)		
\$70,001-\$100,000	9 (8.5)	148 (10.8)	20 (17.5)	177 (11.1)		
≥\$100,001	9 (8.5)	90 (6.6)	7 (6.1)	106 (6.7)		
Level of internet use					177.8 (4)	<.001
Never or rarely	30 (28.3)	34 (2.5)	0 (0.0)	64 (4)		
Several times per week/month	5 (4.7)	70 (5.1)	2 (1.8)	77 (4.8)		
Several times per day	71 (67.0)	1264 (92.4)	112 (98.2)	1447 (91.1)		
Health status					29.4 (4)	<.001
Poor	18 (17.0)	90 (6.6)	4 (3.5)	112 (7.1)		
Normal	51 (48.1)	649 (47.4)	40 (35.1)	740 (46.6)		
Good	37 (34.9)	629 (46)	70 (61.4)	736 (46.3)		
Exercise frequency					11.8 (4)	.019
Never	27 (25.5)	286 (20.9)	19 (16.7)	332 (20.9)		
Sometimes	54 (50.9)	883 (64.5)	72 (63.2)	1009 (63.5)		
Daily	25 (23.6)	199 (14.5)	23 (20.2)	247 (15.6)		

[a] A currency exchange rate of NT $1=US $0.035 is applicable.

[b] Average regular earnings of all employees (including full-time or part-time employees with Taiwanese nationality or foreigners) was New Taiwan dollar (NT) $42,495. The poverty level income ranged from NT$11,648 to NT$24,293 in 2019 in Taiwan.

Table 6 shows that participants at high risk of misjudging health information had a lower education level, lower income, and poor health, and they did not browse or only rarely browsed the internet. These differences were statistically significant. The probability of participants in the medium-risk group having an education level of senior high school, college, or graduate school was 4.65, 13.22, and 18.40 times, respectively, compared with participants in the high-risk group. Participants with an income were more likely to be part of the medium-risk group than those without any income, with an odds ratio (OR) of 1.76 (95% CI 0.58-5.36) to 4.23 (95% CI 1.56-11.47). Participants with good health were more likely to be in the medium-risk group (OR 1.93, 95% CI 0.90-4.14) than those with poor health (OR 2.98, 95% CI 1.30-6.83). Compared with those who were never or only rarely online, those who were online several times every day or several times every week/month were more likely to be in the medium-risk group (OR 3.80, 95% CI 1.71-8.46 and OR 9.19, 95% CI 2.61-32.37, respectively). Similar results were obtained when the high-risk group was compared with the two lower risk groups.

Table 6. Logistic regression analysis across different risk groups (P<.001).

Characteristic	Medium-risk versus high-risk[a], OR[b] (95% CI)	Medium- and low-risk versus high-risk[c], OR (95% CI)
Education level (ref: junior high school or less)[d]		
Senior high school	4.65 (1.54-13.99)	4.50 (1.50-13.52)
University/college	13.22 (4.24-41.23)	13.46 (4.32-41.93)
Graduate school	18.40 (5.03-67.36)	19.11 (5.23-69.90)
Income (New Taiwan $) (ref: $0)[e]		
≤$23,800	0.90 (0.37-2.20)	0.93 (0.38-2.28)
$23,801-$30,000	4.06 (1.44-11.45)	3.86 (1.37-10.87)
$30,001-$50,000	4.08 (1.72-9.69)	3.78 (1.59-8.95)
$50,001-$70,000	4.23 (1.56-11.47)	3.92 (1.45-10.59)
$70,001-$100,000	2.18 (0.75-6.33)	2.16 (0.75-6.23)
≥$100,001	1.76 (0.58-5.36)	1.62 (0.54-4.90)
Level of internet use (ref: never/rarely)		
Several times per week/month	9.19 (2.61-32.37)	9.05 (2.57-31.84)
Several times per day	3.80 (1.71-8.46)	3.95 (1.77-8.79)
Health status (ref: poor)		
Normal	1.93 (0.90-4.14)	2.00 (0.93-4.31)
Good	2.98 (1.30-6.83)	3.22 (1.40-7.39)

[a] –2 log-likelihood: 564.251; Nagelkerke R^2: 0.311.

[b] OR: odds ratio.

[c] –2 log-likelihood: 576.339; Nagelkerke R^2: 0.309.

[d] ref: reference.

[e] A currency exchange rate of NT $1=US $0.035 is applicable.

Discussion

Principal Results

This study developed the DHLA instrument, drawing from the eHEALS, and carried out a reliability and validity analysis. The study also constructed an online health information bank from public sector sources to simulate information likely to be found by real-world users searching for online health information. The study used ROC analysis to find DHLA cutoff values to classify participants into high-, medium-, and low-risk groups based on their likelihood of misinterpreting health information. A higher DHLA score was associated with a lower risk of health information misjudgment. High-risk participants tended to have a low education level, low income, and poor health, and never or rarely browsed the internet for information. The proportions of accurate judgments on moderate and difficult items of health information were lower in the high-risk group (68.9% and 41.5%, respectively) by approximately 10% to 15% compared with the medium-risk (78.9% and 55.0%, respectively) and low-risk (79.8% and 57.0%, respectively) groups. The difference between risk groups was less among the items that were considered easy. The instrument therefore seems to distinguish more accurately between the high-risk and low/medium-risk groups than between the low- and medium-risk groups.

Limitations

The first limitation of this study is that the sample does not represent the population in Taiwan. We did our best to recruit participants across the widest possible range of socioeconomic status, including from among those using social welfare, via colleagues, friends, and relatives. The participants covered different age groups, genders, income levels, and education levels. However, the participants tended to have a high education level, which might have affected their digital health literacy. The second limitation is the data collection mechanism (an online questionnaire platform), which cannot reach participants who are unfamiliar with the internet or mobile devices. However, a previous study found that there were no significant differences in eHEALS scores between paper-based and online questionnaires [31]. In the future, telephone or paper-based channels could be considered for data collection, to broaden the ranges of scores for digital health literacy. This might increase the differences in DHLA scores between the medium- and low-risk groups. The third limitation was the online health information bank, which included 310 items of health information across different difficulty levels (147 easy, 122 moderate, and 41 difficult). This was more items than were used in previous studies [12,18,19]. We considered that the majority of people needed to be able to judge the accuracy of health information from the online health information bank, so we excluded any items regarded as requiring professional knowledge. This may have limited the number of difficult items to 41 and might have made it harder to differentiate between medium- and low-risk groups.

Comparison with Prior Work

The eHEALS is unidimensional, with a Cronbach alpha of 0.88, whereas the DHLA has three dimensions, with a Cronbach alpha of .87. Using a criterion of 0.70 to 0.90 as a measure of good internal consistency [29], the DHLA has adequate internal consistency. Item-scale correlation of the DHLA had a range of 0.27 to 0.87, which was satisfactory because the coefficients were greater than 0.20 [32] for all items. Except for one item on folk remedies (0.27), all the others were over 0.48. The DHLA was developed from the eHEALS, adding two more items and changing the format of response. This gave a different factor structure. Abma et al [33] reported that their relative correlation size of convergent validity ranged from 0.2 to 0.4 with mild to moderate correlation, although the expected direction of the correlation—positive or negative—is also important. DHLA scores showed a negative correlation with age and a positive correlation with education level and level of internet use. The weak correlation coefficient was statistically significant around 0.20, so the convergent validity was considered acceptable.

In a systematic review, Diviani et al [34] found that 10 (26%) of the 38 studies examined the relationship between health literacy and capacity to assess online health information. However, only five studies used the eHEALS as a measure to explore the issues related to online health information searching. The Newest Vital Sign (Pfizer Inc) screening tool has been used in several studies as a performance test to assess the ability to make accurate judgments about online health information [12,17,35]. This tool is based on a nutrition label composed of six questions and requires 3 minutes to complete [36]. A study by Quinn et al [19] developed six health questions as a performance test, including strategic areas such as diabetes, obesity, influenza, nutrition, and analgesic medication. Several studies have used performance tests to collect data, but these take considerable time to complete, often around 1.5 hours [18,37,38]. Studies using performance tests therefore often limit the areas of health information covered, and so do not reflect real-world situations. Taking a long time to collect data might lower the validity of studies because of participants' fatigue. In this study, the majority of participants took approximately 6 minutes to assess five items of health information randomly selected from the online health information bank. The items in the health information bank were drawn from real-world items on the internet. The online health information bank is thus considered to be a reasonable analogy to online health information, but it cannot be regarded as a performance test comparable to those used in previous studies [12,17,18,35,37,38].

The online health information bank contained approximately 300 items, giving it both quantity and variety, which might have resulted in the weak positive correlation coefficients found between difficulty levels and levels of education and internet use (Table 2). Nevertheless, the correlation coefficients were statistically significant, except for easy items. The difficulty levels were categorized as items for which more than 80% (easy), 50% (moderate), and less than 30% (difficult) of people are able to determine credibility. In the study, however, the proportions of correct answers were 86.9% (easy), 78.3% (moderate), and 57.0% (difficult), which was obviously higher than the definitions. This might be because 96.5% of participants had at least a university or college education. However, this study used the health information bank and the DHLA score to distinguish between those at high, medium, and low risk of misinterpreting information by levels of digital health literacy. The study results revealed that the high-risk group had the lowest proportion of correct answers when judging the health information items. This means that the DHLA could be used by health information platforms on the internet as a way to identify—and therefore protect—users at high risk of misinterpreting health information.

Systematic reviews have found that education level is related to digital health literacy [39]. Neter and Brainin [15] studied adults aged 18 years and older and found that digital health literacy was not related to gender or self-reported health, but was associated with age, education level, diseases and conditions, and level of internet use. Xesfingi and Vozikis [40] found that digital health literacy was related to age, education level, and exercise frequency. Del Giudice et al [41] showed that digital health literacy was associated with age, education level, self-reported health, and level of internet health information use. We compared the high-risk group with (1) the medium-risk group, and (2) the combined medium- and low-risk groups, and found that distinguishing factors included education, income, level of internet use, and self-reported health status. Therefore, the factors found in this study are very similar to factors found in previous studies on digital health literacy [15,39-44].

Del Giudice et al [41] found that education level had a weak positive correlation (ρ=0.11, *P*=.001) with digital health literacy. Their sample was divided into two groups: those studying or working in the health sector, and those who were not. Interestingly, in our study, having a health-related job was not statistically significantly correlated with different risk groups. We found that higher education level (ρ=0.22, *P*<.001) was associated with lower risk of health information misjudgment, showing better ability to judge credible health information. This finding suggests that people with a low education level may need help to improve their digital health literacy to avoid misuse of online health information and its potentially negative effects on their health.

Choi and Dinitto [42] demonstrated that people with a low level of education had poor digital health literacy compared with the US population as a whole. A systematic literature review reported an association between income and internet access and digital health literacy [39]. We found that participants with an income of NT $23,800 to $70,000 (US $832.20 to $2447.65) were better able to judge the credibility of internet health information than those with no income. However, the ORs of participants with an income of NT $70,001 (US $2447.68) and NT $100,000 (US $3496.64) or greater was lower than the OR of those with an income of NT $23,801 (US $832.24) to $70,000 (US $2447.65) (2.18 and 1.76, respectively, versus approximately 4; Table 6). This suggests that the relationship between income level and risk group based on DHLA scores may be an inverted U shape. It is possible that those with an income of NT $70,000 (US $832.20) or greater were more cautious in selecting health information and did not want to give incorrect answers, and were thus more likely to respond "unsure," which resulted in a lower probability of being in the medium- and low-risk groups.

Studies have shown that self-reported health status is significantly correlated with digital health literacy [41,45,46]. We found that participants with poor health were more likely to be in the high-risk group, with low digital health literacy. It is therefore important to help those with poor health status to improve their health literacy through health education. This will reduce their risk of using false information from the internet.

Zrubka et al [44] indicated that many studies found an association between digital health literacy and level of internet use, although some studies have found that the correlation is only weak [18]. We found that participants who go online several times per day, per week, or per month were better able to identify correct health information and were at a lower risk of health information misjudgment than those who rarely or never use the internet. However, those who used the internet several times per week or month had the highest probability of being in the medium- or low-risk groups (OR 9.19, 95% CI 2.61-32.37), followed by those using the internet several times per day (OR 3.80, 95% CI 1.71-8.46), compared with those who never or rarely used the internet. It seems likely that those who browse the internet several times per day might become blasé about the quality of the information available, whereas those who use the internet several times per week or month may be more careful. It may be helpful to improve internet access for people who never or rarely use the internet and encourage them to use it more often. However, websites may also need to provide a mechanism to protect them from misinterpreting information.

Conclusions

Our results showed that DHLA scores could distinguish between those at low, medium, and high risk of misjudging health information on the internet. Health information platforms on the internet could therefore use the DHLA to set up mechanisms to protect users from misusing health information and thereby avoid harming their health. This would mean incorporating the DHLA into the website and verifying it with a larger sample. Simultaneously, this approach may decrease the possibility that users will receive erroneous health information that could threaten their quality of life.

Acknowledgments

We thank all the respondents for their participation in the study. We also thank Melissa Leffler, MBA, from Edanz Group Japan Inc, for editing a draft of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Digital health literacy assessment.
[DOCX File , 23 KB - jmir_v22i12e19767_app1.docx]

Multimedia Appendix 2
Sample items of online health information bank.
[DOCX File , 21 KB - jmir_v22i12e19767_app2.docx]

References

1. ITU. Measuring digital development: Facts and figures 2019. Geneva, Switzerland: International Telecommunication Union; 2019.
2. European Commission. European Citizens' Digital Health Literacy: Report. Brussels: European Commission; 2014.
3. HINTS response: The most recent time you looked for information about health or medical topics, where did you go first? Health Information National Trends Survey 5 (HINTS 5): Cycyle 3 web site. National Cancer Institute. URL: https://hints.cancer.gov/view-questions-topics/question-details.aspx?PK_Cycle=12&qid=688 [accessed 2020-06-04]
4. National Development Council. 2016 Digital Opportunity Survey for Individuals and Households (AE010014). Taipei: National Development Council; 2017.
5. Song H, Omori K, Kim J, Tenzek KE, Morey HJ, Lin W, et al. Trusting Social Media as a Source of Health Information: Online Surveys Comparing the United States, Korea, and Hong Kong. J Med Internet Res 2016 Mar 14;18(3):e25 [FREE Full text] [doi: 10.2196/jmir.4193] [Medline: 26976273]
6. Lee HY, Wang CC, Lo SK, Fang WC. An exploratory study on the transnational dissemination phenomenon of internet rumors. The Journal of Information Society 2003 Jul;5:175-208 [FREE Full text]
7. National Development Council. 2019 digital opportunity survey for individuals and households. Taipei: National Development Council; Mar 20, 2020.
8. DiFonzo N, Robinson NM, Suls JM, Rini C. Rumors about cancer: content, sources, coping, transmission, and belief. J Health Commun 2012;17(9):1099-1115. [doi: 10.1080/10810730.2012.665417] [Medline: 22724591]
9. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
10. Koo M, Norman CD, Chang H. Psychometric evaluation of a Chinese version of the eHealth Literacy Scale (eHEALS) in school age children. Glob J Health Educ Promot 2012 Jan 1;15(1):29-36.
11. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowledge management & e-learning 2015 Dec 15;7(4):576-600. [doi: 10.34105/j.kmel.2015.07.038]
12. Ghaddar SF, Valerio MA, Garcia CM, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health 2012 Jan;82(1):28-36. [doi: 10.1111/j.1746-1561.2011.00664.x] [Medline: 22142172]
13. MedlinePlus. URL: https://medlineplus.gov/ [accessed 2020-11-13]
14. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]
15. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
16. Diviani N, van DPB, Meppelink CS, van WJCM. Exploring the role of health literacy in the evaluation of online health information: Insights from a mixed-methods study. Patient Educ Couns 2016 Jun;99(6):1017-1025. [doi: 10.1016/j.pec.2016.01.007] [Medline: 26817407]
17. Noblin AM, Rutherford A. Impact of Health Literacy on Senior Citizen Engagement in Health Care IT Usage. Gerontol Geriatr Med 2017;3:2333721417706300 [FREE Full text] [doi: 10.1177/2333721417706300] [Medline: 28508022]
18. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
19. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. Computers in Human Behavior 2017 Apr;69:256-267. [doi: 10.1016/j.chb.2016.12.032]
20. Ma Z, Wu M. The Psychometric Properties of the Chinese eHealth Literacy Scale (C-eHEALS) in a Chinese Rural Population: Cross-Sectional Validation Study. J Med Internet Res 2019 Oct 22;21(10):e15720 [FREE Full text] [doi: 10.2196/15720] [Medline: 31642811]
21. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the Measurement Properties of the eHealth Literacy Scale (eHEALS) Among Baby Boomers: A Multinational Test of Measurement Invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
22. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
23. Wu CC. An analysis of health-related questions on social Q&A. Taipei: National Taiwan Normal University; Jan 2014.
24. Fact check column. Taiwan Food and Drug Administration Web site. URL: https://www.fda.gov.tw/TC/news.aspx?cid=5049 [accessed 2020-11-13]
25. Healthcare fact check column. Health Promotion Administration Web site. URL: https://www.hpa.gov.tw/Pages/List.aspx?nodeid=70 [accessed 2020-11-13]
26. Healthcare fact check column dataset. Health Promotion Administration. Government Open Data Web site. URL: https://data.gov.tw/dataset/43338 [accessed 2020-11-13]

27. Fact check column dataset about Foods and Drugs. Government Open Data Web site. Taiwan Food and Drug Administration. URL: https://data.gov.tw/dataset/17148 [accessed 2020-11-13]
28. Hair Jr JF, Hult GTM, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Thousand Oaks, California: Sage Publications Inc; 2017.
29. Terwee CB, Bot SDM, de BMR, van DWDAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
30. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochem Med (Zagreb) 2016 Oct 15;26(3):297-307 [FREE Full text] [doi: 10.11613/BM.2016.034] [Medline: 27812299]
31. Juvalta S, Kerry MJ, Jaks R, Baumann I, Dratva J. Electronic Health Literacy in Swiss-German Parents: Cross-Sectional Study of eHealth Literacy Scale Unidimensionality. J Med Internet Res 2020 Mar 13;22(3):e14492 [FREE Full text] [doi: 10.2196/14492] [Medline: 32167476]
32. Wang M, Batt K, Kessler C, Neff A, Iyer NN, Cooper DL, et al. Internal consistency and item-total correlation of patient-reported outcome instruments and hemophilia joint health score v2.1 in US adult people with hemophilia: results from the Pain, Functional Impairment, and Quality of life (P-FiQ) study. Patient Prefer Adherence 2017;11:1831-1839 [FREE Full text] [doi: 10.2147/PPA.S141391] [Medline: 29123383]
33. Abma IL, Rovers M, van der Wees PJ. Appraising convergent validity of patient-reported outcome measures in systematic reviews: constructing hypotheses and interpreting outcomes. BMC Res Notes 2016 Apr 19;9:226 [FREE Full text] [doi: 10.1186/s13104-016-2034-2] [Medline: 27094345]
34. Diviani N, van DPB, Giani S, van WJC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
35. Monkman H, Kushniruk AW, Barnett J, Borycki EM, Greiner LE, Sheets D. Are Health Literacy and eHealth Literacy the Same or Different? Stud Health Technol Inform 2017;245:178-182. [Medline: 29295077]
36. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
37. Feufel MA, Stahl SF. What do web-use skill differences imply for online health information searches? J Med Internet Res 2012;14(3):e87 [FREE Full text] [doi: 10.2196/jmir.2051] [Medline: 22695686]
38. van Deursen AJAM, van Dijk JAGM. Internet skills performance tests: are people ready for eHealth? J Med Internet Res 2011 Apr 29;13(2):e35 [FREE Full text] [doi: 10.2196/jmir.1581] [Medline: 21531690]
39. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: A systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
40. Xesfingi S, Vozikis A. eHealth Literacy: In the Quest of the Contributing Factors. Interact J Med Res 2016 May 25;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]
41. Del Giudice P, Bravo G, Poletto M, De Odorico A, Conte A, Brunelli L, et al. Correlation Between eHealth Literacy and Health Literacy Using the eHealth Literacy Scale and Real-Life Experiences in the Health Sector as a Proxy Measure of Functional Health Literacy: Cross-Sectional Web-Based Survey. J Med Internet Res 2018 Oct 31;20(10):e281 [FREE Full text] [doi: 10.2196/jmir.9401] [Medline: 30381283]
42. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
43. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
44. Zrubka Z, Hajdu O, Rencz F, Baji P, Gulácsi L, Péntek M. Psychometric properties of the Hungarian version of the eHealth Literacy Scale. Eur J Health Econ 2019 Jul;20(Suppl 1):57-69 [FREE Full text] [doi: 10.1007/s10198-019-01062-1] [Medline: 31098883]
45. De Caro W, Corvo E, Marucci AR, Mitello L, Lancia L, Sansoni J. eHealth Literacy Scale: An Nursing Analisys and Italian Validation. Stud Health Technol Inform 2016;225:949. [Medline: 27332425]
46. Paramio Pérez G, Almagro BJ, Hernando Gómez Á, Aguaded Gómez JI. [Validation of the eHealth Literacy Scale (eHEALS) in Spanish University Students]. Rev Esp Salud Publica 2015;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]

Abbreviations

ANOVA: analysis of variance
DHLA: digital health literacy assessment
eHEALS: eHealth Literacy Scale

NT: New Taiwan dollar
OR: odds ratio
ROC: receiver operating characteristic

Edited by G Eysenbach; submitted 30.04.20; peer-reviewed by Z Ma, Z Zrubka, P Schulz, S Ghaddar; comments to author 25.05.20; revised version received 16.07.20; accepted 26.10.20; published 21.12.20.

Please cite as:
Liu P, Yeh LL, Wang JY, Lee ST
Relationship Between Levels of Digital Health Literacy Based on the Taiwan Digital Health Literacy Assessment and Accurate Assessment of Online Health Information: Cross-Sectional Questionnaire Study
J Med Internet Res 2020;22(12):e19767
URL: http://www.jmir.org/2020/12/e19767/
doi:10.2196/19767
PMID:33106226

Original Paper

Transformation of Pathology Reports Into the Common Data Model With Oncology Module: Use Case for Colon Cancer

Borim Ryu[1], MSc; Eunsil Yoon[1], MSc; Seok Kim[1], MPH; Sejoon Lee[2], PhD; Hyunyoung Baek[1], MSc; Soyoung Yi[1], BSc; Hee Young Na[2], BSc; Ji-Won Kim[3], MD, PhD; Rong-Min Baek[4], MD, PhD; Hee Hwang[1], MD; Sooyoung Yoo[1], PhD

[1]Office of eHealth Research and Business, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[2]Department of Pathology and Translational Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[3]Division of Hematology and Medical Oncology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[4]Department of Plastic Surgery, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

Corresponding Author:
Sooyoung Yoo, PhD
Office of eHealth Research and Business
Seoul National University Bundang Hospital
82 173rd Street, Gumi-ro, Bundang-gu
Seongnam, 13620
Republic of Korea
Phone: 82 317878980
Email: yoosoo0@gmail.com

Abstract

Background: Common data models (CDMs) help standardize electronic health record data and facilitate outcome analysis for observational and longitudinal research. An analysis of pathology reports is required to establish fundamental information infrastructure for data-driven colon cancer research. The Observational Medical Outcomes Partnership (OMOP) CDM is used in distributed research networks for clinical data; however, it requires conversion of free text–based pathology reports into the CDM's format. There are few use cases of representing cancer data in CDM.

Objective: In this study, we aimed to construct a CDM database of colon cancer–related pathology with natural language processing (NLP) for a research platform that can utilize both clinical and omics data. The essential text entities from the pathology reports are extracted, standardized, and converted to the OMOP CDM format in order to utilize the pathology data in cancer research.

Methods: We extracted clinical text entities, mapped them to the standard concepts in the Observational Health Data Sciences and Informatics vocabularies, and built databases and defined relations for the CDM tables. Major clinical entities were extracted through NLP on pathology reports of surgical specimens, immunohistochemical studies, and molecular studies of colon cancer patients at a tertiary general hospital in South Korea. Items were extracted from each report using regular expressions in Python. Unstructured data, such as text that does not have a pattern, were handled with expert advice by adding regular expression rules. Our own dictionary was used for normalization and standardization to deal with biomarker and gene names and other ungrammatical expressions. The extracted clinical and genetic information was mapped to the Logical Observation Identifiers Names and Codes databases and the Systematized Nomenclature of Medicine (SNOMED) standard terminologies recommended by the OMOP CDM. The database-table relationships were newly defined through SNOMED standard terminology concepts. The standardized data were inserted into the CDM tables. For evaluation, 100 reports were randomly selected and independently annotated by a medical informatics expert and a nurse.

Results: We examined and standardized 1848 immunohistochemical study reports, 3890 molecular study reports, and 12,352 pathology reports of surgical specimens (from 2017 to 2018). The constructed and updated database contained the following extracted colorectal entities: (1) NOTE_NLP, (2) MEASUREMENT, (3) CONDITION_OCCURRENCE, (4) SPECIMEN, and (5) FACT_RELATIONSHIP of specimen with condition and measurement.

Conclusions: This study aimed to prepare CDM data for a research platform to take advantage of all omics clinical and patient data at Seoul National University Bundang Hospital for colon cancer pathology. A more sophisticated preparation of the pathology data is needed for further research on cancer genomics, and various types of text narratives are the next target for additional research on the use of data in the CDM.

(J Med Internet Res 2020;22(12):e18526) doi:10.2196/18526

KEYWORDS

common data model; natural language processing; oncology module; colon cancer; electronic health record; oncology; pathology; clinical data

Introduction

Colorectal cancer is the third most common cancer in the world after lung cancer and breast cancer, and the second most common cause of cancer deaths in the world after lung cancer [1]. In addition, the incidence of colorectal cancer in Korea is continuously increasing owing to the westernization of diet and the widespread use of colonoscopy [2]. To determine treatment and prognosis, clinical and pathologic staging are both crucial. Pathology reports vary in format worldwide. The heterogeneity of pathology reports is not unique to colorectal cancer; there has been a growing need to standardize pathology reports [3]. Pathologic diagnosis is based on gross examination, microscopic examination, and sometimes molecular testing. Although some hospitals report molecular testing results in a structured format, biomarker results in the pathology report are usually recorded as unstructured free text or template-based semistructured text in electronic health record (EHR) systems. This unstructured document must be converted into a structure that can be processed by a computer.

Unstructured clinical narratives may summarize patients' medical history, diagnoses, medications, immunizations, allergies, radiology images, and laboratory test results in the form of progress notes, discharge reports, etc, providing a valuable resource for computational phenotyping [4]. Previous studies have applied controlled vocabularies such as the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) collection to recognize various expressions of the same medical concepts in pathology and used the Unified Medical Language System (UMLS) as a metathesaurus [5-7]. Other significant efforts have been devoted to the implementation of open-source, standards-based systems to improve the portability of EHR-based phenotype definitions (eg, eMERGE [Electronic Medical Records and Genomics] [8] and PhEMA [Phenotype Execution and Modeling Architecture] [9]).

Common data models (CDMs) are healthcare data models with a standard structure. An example is the Observational Medical Outcomes Partnership (OMOP) CDM, adopted and maintained by Observational Health Data Sciences and Informatics (OHDSI) [10-13]. OHDSI is an open-science community that aims to improve health by empowering the community to collaboratively generate evidence that promotes better health decisions and care [14]. OHDSI conducts methodological research to establish scientific best practices for the appropriate use of OMOP CDM data and develops open-source analytics software for research use. The OHDSI oncology working group incorporated fundamental structural and semantic support into the OMOP CDM to represent clinical cancer disease and treatment data, significantly improving the specificity of cancer cohort definitions. To represent cancer diagnoses using the combination of histology and topography in the OMOP CDM Condition domain (Condition_Occurrence), without changes to the existing structure, they have proposed a precoordinated collapse of the International Classification of Diseases for Oncology (ICD-O) axes—histology and topography—to a single OMOP-originated concept representing a unique cancer diagnosis, which preserves linkages between these single codes and the ICD-O axes in the OMOP standard vocabulary.

To transform EHR data into the OMOP CDM format, data formalization and vocabulary mapping should be performed in advance. Thus, most structured clinical EHR data, such as diagnosis, medications, lab tests, and vital signs, are usually converted to the CDM first. However, a large amount of data in EHRs—including pathology reports—is recorded and stored in an unstructured or semistructured form. In particular, it is essential to extract and refine major clinical entities from the pathology reports through natural language processing (NLP) to utilize both clinical and genomics data in CDM-based cancer research.

In this study, we transformed and incorporated the colorectal cancer pathology reports of a tertiary general university hospital into the OMOP CDM format by developing an NLP module. The primary objective of this research is to (1) extract major biomarker entities from the pathology report, (2) convert them into the OMOP CDM format by mapping the vocabulary to standardized terminology, and (3) demonstrate the ability of the OMOP CDM to represent biomarker data using the OMOP CDM oncology extension module.

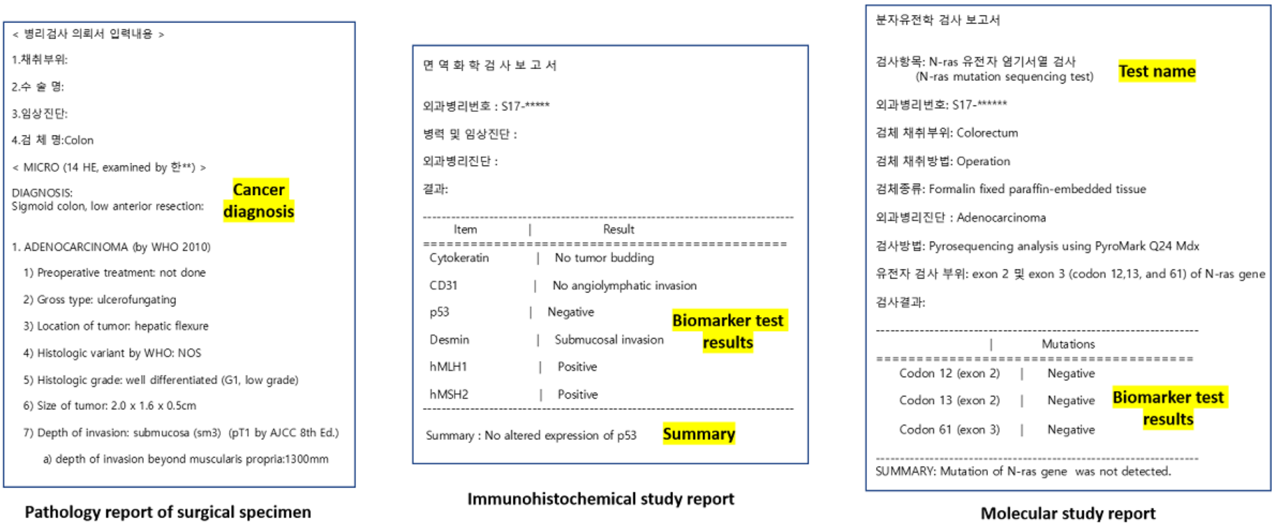
Methods

Study Data

The study site, Seoul National University Bundang Hospital (SNUBH), has converted EHR data from a 15-year period—from the opening of SNUBH until December 2018—into the OMOP CDM format. The CDM database contains data on more than 1.7 million patients. Although patient data such as basic demographic information, medical history, family history, diagnosis, drug exposures, test results, vital signs, surgeries, and procedures were converted into the OMOP CDM standard, free-text results, such as pathology reports, are stored in the form of free text in a table called NOTE. It is necessary to extract these data from the reports through text processing techniques for use in clinical research. In this study, the pathology reports of patients diagnosed with colorectal cancer from 2017 to 2018 were processed to investigate the ability of the OMOP CDM Oncology module to represent three types of pathology reports: pathology reports of surgical specimens, immunohistochemical study reports, and molecular study reports. This study was approved for exemption by the SNUBH Institutional Review Board.

The original data were derived from the NOTE table of the SNUBH OMOP CDM database and included the original free-text data of pathology reports of surgical specimens, immunohistochemical study reports, and molecular study reports (Figure 1). Surgical number, biomarker name, test result, and summary information were extracted from the immunohistochemical study tests. In pathology reports, a molecular study is a method of analyzing protein expression in cells; in this study, surgical number, test name, gene name, mutation examination result, and summary information were extracted from molecular study tests. The surgical pathology report is an essential part of patient care because it documents the pathologic findings in tissues removed from patients for diagnostic or therapeutic reasons [15]. Diagnosis names were extracted from pathology reports of surgical specimens.

Figure 1. Contents of information to extract by report type.



```
< 병리검사 의뢰서 입력내용 >
1.채취부위:
2.수 술 명:
3.임상진단:
4.검 체 명:Colon
< MICRO (14 HE, examined by 한**) >
DIAGNOSIS:
Sigmoid colon, low anterior resection:

1. ADENOCARCINOMA (by WHO 2010)
  1) Preoperative treatment: not done
  2) Gross type: ulcerofungating
  3) Location of tumor: hepatic flexure
  4) Histologic variant by WHO: NOS
  5) Histologic grade: well differentiated (G1, low grade)
  6) Size of tumor: 2.0 x 1.6 x 0.5cm
  7) Depth of invasion: submucosa (sm3)  (pT1 by AJCC 8th Ed.)
      a) depth of invasion beyond muscularis propria:1300mm
```

Cancer diagnosis

Pathology report of surgical specimen

```
면 역 화 학 검 사 보 고 서

외과병리번호 : S17-*****
병력 및 임상진단 :
외과병리진단 :
결과:
--------------------------------------------------------------
      Item          |              Result
==============================================================
 Cytokeratin          |   No tumor budding
 CD31                 |   No angiolymphatic invasion
 p53                  |   Negative
 Desmin               |   Submucosal invasion
 hMLH1                |    Positive
 hMSH2                |    Positive
--------------------------------------------------------------
Summary : No altered expression of p53
```

Biomarker test results

Summary

Immunohistochemical study report

```
분자유전학 검사 보고서

검사항목: N-ras 유전자 염기서열 검사
        (N-ras mutation sequencing test)
외과병리번호: S17-******
검체 채취부위: Colorectum
검체 채취방법: Operation
검체종류: Formalin fixed paraffin-embedded tissue
외과병리진단 : Adenocarcinoma
검사방법: Pyrosequencing analysis using PyroMark Q24 Mdx
유전자 검사 부위: exon 2 및 exon 3 (codon 12,13, and 61) of N-ras gene
검사결과:

-----------------------------------------------------
                     |        Mutations
=====================================================
    Codon 12 (exon 2)    |     Negative
    Codon 13 (exon 2)    |     Negative
    Codon 61 (exon 3)    |     Negative

-----------------------------------------------------
SUMMARY: Mutation of N-ras gene  was not detected.
```

Test name

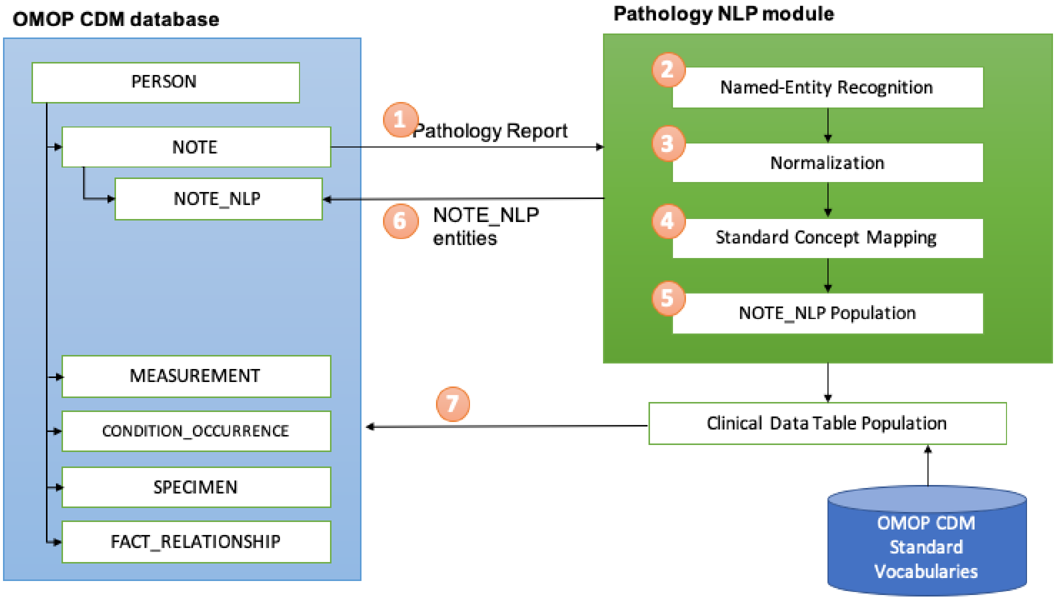
Biomarker test results

Molecular study report

Processing Model for Rule-Based Text Extraction

The types of key information recorded in pathology reports vary based on the pathological tests conducted. We first reviewed the format and text entities of the extracted original report and selected the target entities and values for information extraction. Rule-based text processing based on regular expressions was developed to extract the clinical entities of the test subject and result information depending on the type of report. In the immunohistochemical study reports and molecular study reports, target gene or protein information and the test result values were recognized and extracted, whereas pathology reports of surgical specimens were processed for information on the diagnosis and the region that was operated on. This rule-based approach is ideally suited for the retrieval of unstructured pathologic clinical entities and test result values in the SNUBH EHR notes. Figure 2 shows the NLP module developed in this study and depicts the overall study process, with the numbers in the circles indicating the sequence of steps.

Figure 2. Overall process of this study. CDM: common data model. NLP: natural language processing. OMOP: Observational Medical Outcomes Partnership.



The measurement value extraction system was packaged through a regular expression–based Python NLP pipeline using a method similar to that employed in previous research [16]. Our system recognizes clinical entities such as the biomarkers in a test and their values. Our system performs three broad steps: (1) it identifies the measurement name, such as gene or protein name, in the text; (2) it identifies the measurement value in the text; and (3) it transforms the appropriate value expressions according to their relationship. For immunohistochemical and molecular study tests, rules were created for extracting the target gene name, protein name information, and test result value. For example, surgical number, biomarker name, test result, and summary information were extracted from the immunohistochemical study tests (example in Figure 1). Text entities for a test result could include expressions such as “EGFR (GI)” (epidermal growth factor receptor; gastrointestinal) and “1+/3.” The term “EGFR (GI)” is regularized into “EGFR” and “epidermal growth factor receptor Ag [presence] in tissue by immune stain”. Through this clinical entity recognition, we could structure data into tables that can be transformed into the NOTE_NLP table after term standardization in the CDM (Table 1 and Figure 1).

Several modifiers were added to enter data extracted from the NOTE_NLP table. For the immunohistochemical study report and the molecular study test report, the raw text of the test results was entered in the value_as_narrative column, and the information on whether the normalized value was a negation expression and the concept_id were included in the value_as_concept_id column. The names of each report were added to the section_source_value column, and inspection item information was included for molecular study test reports. In addition, the surgical pathology number was added to the sub_id column to be linked to surgical specimen pathology reports.

We used Python version 3.6 and regular expressions to process three types of colon cancer pathology reports according to each report’s characteristics. For example, the specimen and diagnosis names were extracted from surgical specimen pathology reports, with the specimen name written in each report as “SAMPLE NAME: colon,” using a regular expression to extract the items after “SAMPLE NAME:.” Likewise, the diagnosis name is found after the entity expression “DIAGNOSIS:.” From the immunohistochemical study report, items following “SUMMARY:” and prior to “EXAMINER:” were extracted as summary information. In addition, biomarker results from the immunohistochemical study and molecular study pathology reports were extracted in a semistructured text table with continuous values such as hyphens (-). To deal with this text, data fields were separated based on the continuous value of hyphen (-), with biomarker name on the left of the vertical bar (|) and result value on the right. Refer to Multimedia Appendix 1 for processing rules (Table S1).

Table 1. NOTE_NLP table for pathology report data.

NOTE_NLP field	Surgical pathology report	Immunochemistry report	Molecular study report
NOTE_NLP_ID	12345	23456	34567
NOTE_ID	1	2	3
SNIPPET	ductal adenocarcinoma, moderately differentiated > poorly differentiated	EGFR (GI) | 1+/3	Codon 12 (exon 2) | Positive
OFFSET	560,632	170,188	200,231
SECTION_CONCEPT_ID	3025891 (Pathology report final diagnosis narrative)	40758358 (Immune stain study)	3001274 (NRAS gene mutations found [identifier] in blood or tissue by molecular genetics method nominal)
LEXICAL_VARIANT	Adenocarcinoma	EGFR	Codon 12 (exon 2)
NOTE_NLP_CONCEPT_ID	0	3016231 (Epidermal growth factor receptor Ag [presence] in tissue by immune stain)	0
NOTE_NLP_SOURCE_CONCEPT_ID	44498791 (Tubular adenoma, NOS[a])	0	0
TERM_MODIFIERS	Negated=FALSE; sub_id=S 110023456; section_source_value=Pathology reports of surgical specimen	Negated=FALSE; value_as_concept_id=9191; value_as_narrative=1+/3; sub_id=S 120034567; section_source_value=Immunochemistry test report	Negated=FALSE; value_as_concept_id=9191; value_as_narrative=Missense mutation [c.38G>A, p.Gly14Asp]; sub_id=S 130045678; section_source_value=N-ras Gene sequencing test

[a] NOS: not otherwise specified.

Vocabulary Standardization for Pathology Reports

To standardize the extracted pathologic data, protein or gene laboratory tests and pertinent test result values were mapped into the OMOP standard concept vocabulary and reviewed by a clinician, a nurse, and a bioinformatics expert. The test names were mapped to standard concepts of Logical Observation Identifiers Names and Codes (LOINC) in the MEASUREMENT domain, and the result values were normalized to those concepts (eg, categorized as positive or negative).

According to the OMOP CDM Oncology module, information on the cancer diagnosis and the region operated on, taken from the surgical specimen pathology reports, was mapped to the ICD-O, Third Edition, and then stored in the CONDITION_OCCURRENCE table with ICD-O-3 and its mapped SNOMED CT codes.

To map the extracted clinical and gene information to a standard concept_id, we constructed a dictionary table for the test names and result values that appear in the free-text pathology reports. For example, in immunohistochemical study reports, the test name for EGFR is concept_id 3016231 (Epidermal growth factor receptor Ag [presence] in tissue by immune stain), which is the LOINC code 32581-1 with a positive or negative result value. The result value of “1+/3” was normalized and mapped to the concept_id of 10828004 (Positive), which is SNOMED CT code 9191. The molecular study test report comprised results of four types of examinations: *NRAS* (OMIM 164790) mutation, *KRAS* (OMIM 190070) mutation, *BRAF* (OMIM 164757) mutation, and microsatellite instability. For each test result value, the normalization rules for test result values were separately defined and coded, as shown in Table 2.

Table 2. Defined mapping rules for OMOP CDM concept_id, LOINC code, and test results from the molecular study pathology reports.

Test name	CONCEPT_ID	LOINC code	RESULT_Attribute	Normalization rules for results value concept
NRAS mutation	3001274	21719-0	Codon 12 (exon 2), Codon 13 (exon 2), Codon 61 (exon 3)	If 1 or more positive values are present, treat as total positive
KRAS mutation	36203353	85509-8	Codon 12 (exon 2), Codon 13 (exon 2), Codon 61 (exon 3)	If 1 or more positive values are present, treat as total positive
MSI[a]	3047348	43368-0	BAT26, BAT25, D5S346, D17S250, D2S123	Categorized as MSI-H (high) if more than 1 value is positive, MSI-L (low) if 1 value is positive, and MSS[b] if no positive value is present
BRAF mutation	40761583	58483-9	V600E, K601E	If more than 1 positive value is present, treat as total positive

[a] MSI: microsatellite instability.

[b] MSS: microsatellite stable.

A total of 78 biomarkers were used in immunohistochemical reports, most of which are measurement domains, and the concept_id for each was retrieved and mapped from the OMOP CDM. However, there were 3 undetected biomarkers that were used with their own concept_id (3/78, 4%). Refer to Multimedia Appendix 2 for a list of mapped concepts (Table S2).

After term mapping, the NOTE_NLP table of the CDM was derived. To integrate the extracted data into the existing CDM data, the main results of immunohistochemical and molecular study tests were updated in the MEASUREMENT table. In addition, for the cancer diagnosis from surgical specimen pathology reports, relevant information was added to the CONDITION_OCCURRENCE table. To define the relationship between the MEASUREMENT, CONDITION_OCCURRENCE, and SPECIMEN tables, we populated the CDM FACT_RELATIONSHIP table. With the two existing concepts, “Specimen source identity (SNOMED)” and “Has specimen (SNOMED),” we could define the target specimen and its attribute.

System Validation

Two experts manually reviewed the transformed data in the NOTE_NLP table in 100 randomly selected pathology reports. To verify the accuracy of the extracted values for each table item, the nurse and medical information specialist manually checked the original pathology document and extracted text values. Furthermore, two reviewers manually reviewed the original document and checked the results of the processing rules.

Results

Major Text-Entity Extraction and Mapping Onto International Standard Terminology

Pathologic clinical information on colorectal cancer was extracted through NLP of immunohistochemical studies, molecular studies, and surgical specimen pathology reports to construct a CDM-formatted NOTE_NLP table. NLP extraction and terminology standardization were performed for 1848 immunohistochemical study reports, 3890 molecular study reports, and 12,352 surgical specimen pathology reports. The number of items, such as protein or gene name values, extracted from the immunohistochemical study reports was 6092, and the number extracted from the molecular study reports was 13,953; the summarized essential information was then updated in the existing MEASUREMENT table. Cancer diagnosis entities from the surgical specimen pathology reports (25,902) were delivered to the CONDITION_OCCURRENCE table. Figure 3 shows the number of each biomarker extracted from molecular study pathology reports by result, and Figure 4 shows the frequency of each diagnosis name and its surgical regions as a heatmap plot. *KRAS* mutations, *NRAS* mutations, *BRAF* mutations, and microsatellite instability are tested in the hospital. For each test result value, the normalization rules for test result values were separately defined and coded as described in Table 2. The majority of the test results are “negative” or “microsatellite instability high.” In Table 3, we included evaluation results for each attribute extracted and transformed into standard terminology concepts for each pathology report. The results of the attribute value extraction task are quite good. The attributes derived from surgical specimen pathology and molecular study reports were recognized 100% correctly among randomly selected test reports. “No endolymphatic tumor emboli” or “no lymphoepithelial lesion” values were recognized as positive in a few immunohistochemical study entities. This is because we have created rules that focus on processing the term “negative” when we process biomarker text.

According to Figure 4, there were 11 colon surgical regions at SNUBH from 2017 to 2018, and the most frequent diagnosis names were “tubular adenoma, nos” (not otherwise specified) and “adenocarcinoma, nos.” Positive and negative biomarker results extracted from immunohistochemical study reports are shown in Table 4.

Figure 3. Each biomarker and its results extracted from molecular study pathology reports.

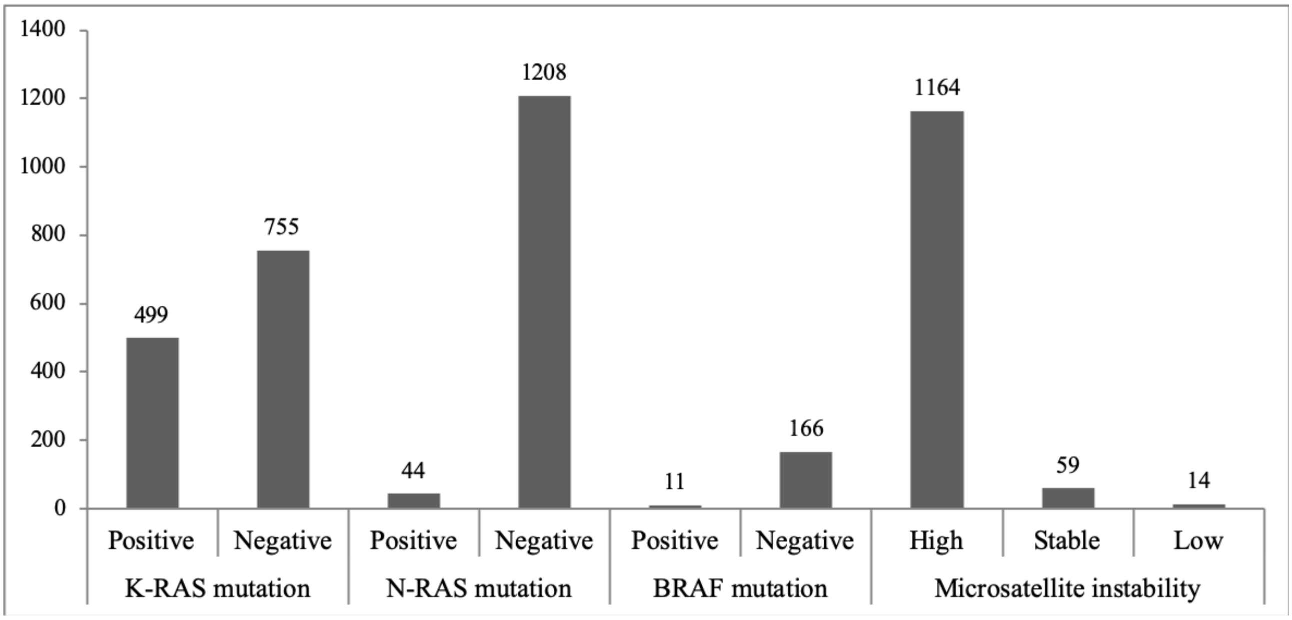


Figure 4. Frequency distribution in surgical pathology reports. NOS: not otherwise specified.

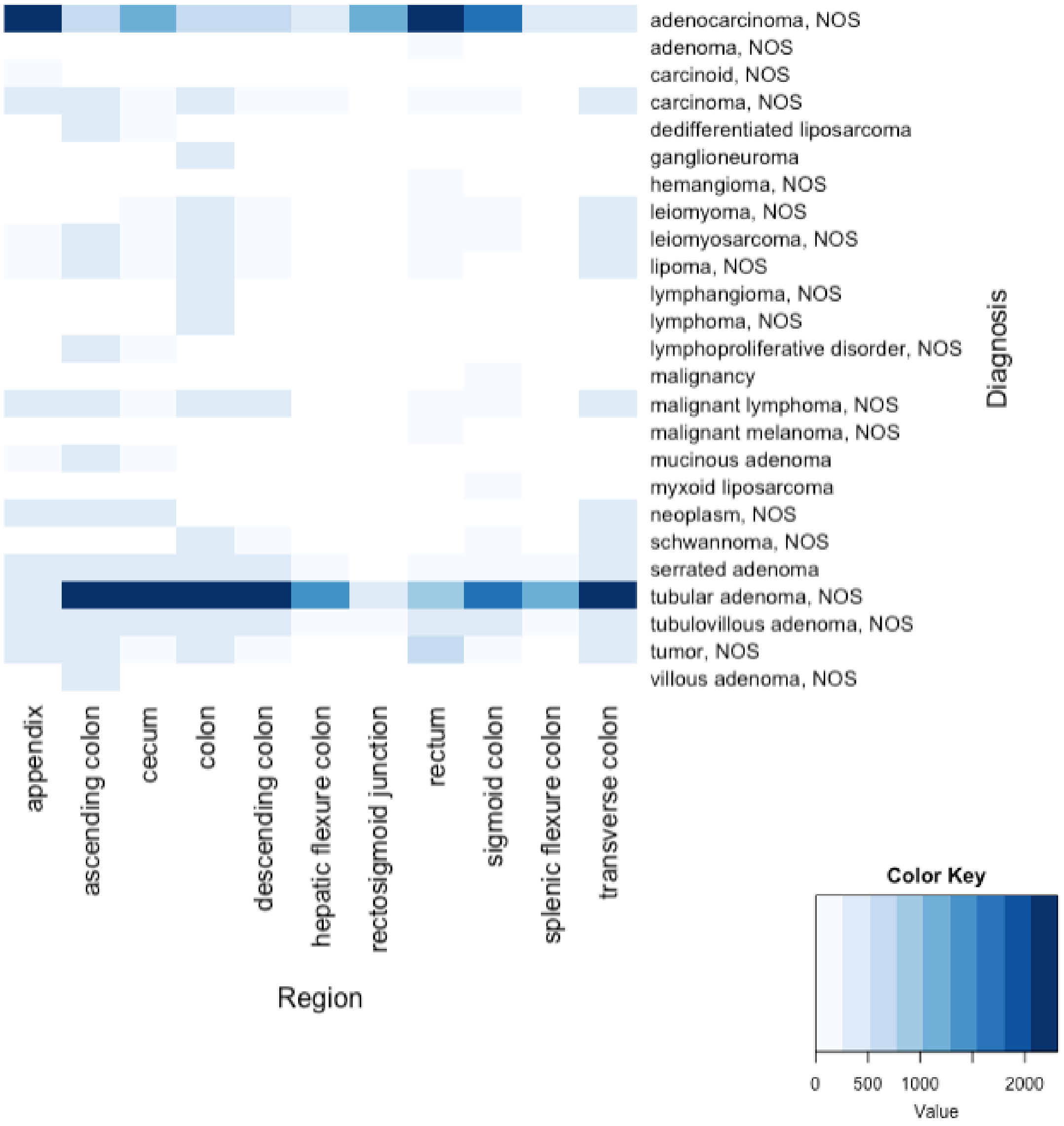


Table 3. Evaluation results for attributes found in 100 pathology test reports.

Type of report	Attribute example	Entities, N	Precision, n (%)	Recall, n (%)	F-measure
Surgical specimen pathology report	ductal adenocarcinoma, moderately differentiated > poorly differentiated	109	109 (100)	109 (100)	100
Immunohistochemical study	EGFR (GI) | 1+/3	367	362 (98.6)	367 (100)	99.31
Molecular study	Codon 12 (exon 2) | Positive	100	100 (100)	100 (100)	100

Table 4. Top 20 biomarker results from immunohistochemistry tests.

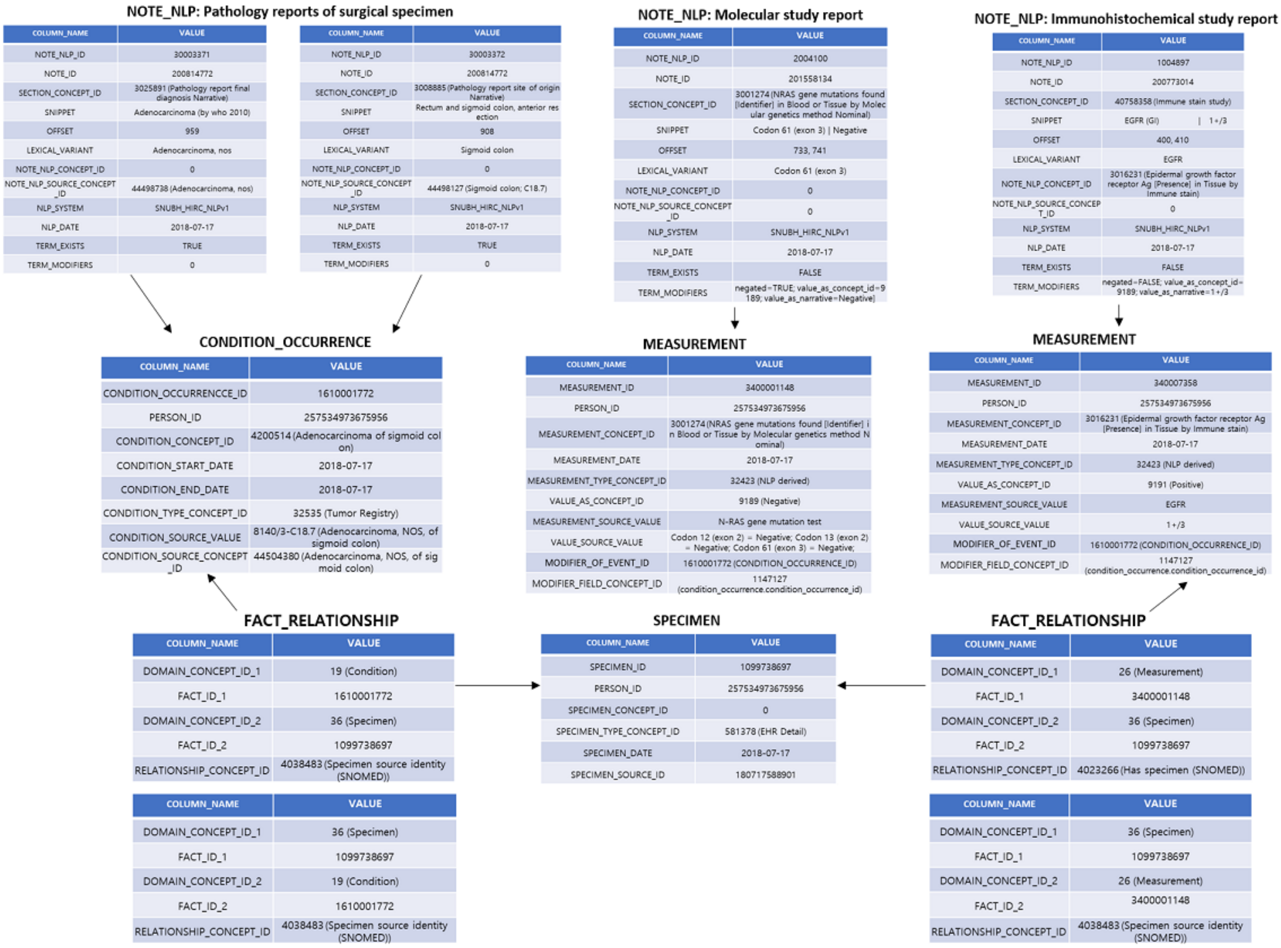
CONCEPT_ID	CONCEPT_NAME	Positive (N=4626), n (%)	Negative (N=2096), n (%)
3016231	EGFR	1360 (29.4)	33 (1.6)
21493968	hMLH1	361 (7.8)	25 (1.2)
3046605	Ki-67	359 (7.8)	0 (0.0)
21493983	PTEN	261 (5.6)	76 (3.6)
3017031	p53	111 (2.4)	212 (10.1)
21493982	BRAF	12 (0.3)	310 (14.8)
3027870	CD3	153 (3.3)	160 (7.6)
3019066	C-erbB2	75 (1.6)	194 (9.3)
3026213	CD20	171 (3.7)	37 (1.8)
21493969	hMSH2	185 (4.0)	18 (0.9)
3051327	bcl-6	100 (2.2)	56 (2.7)
21493970	hMSH6	137 (3.0)	18 (0.9)
21493971	PMS2	133 (2.9)	22 (1.1)
21492142	Cyclin D1	0 (0.0)	151 (7.2)
3002495	Desmin	132 (2.9)	18 (0.9)
3052827	CD8	140 (3.0)	8 (0.4)
3041284	CD10	86 (1.9)	60 (2.9)
3040360	Cytokeratin	79 (1.7)	67 (3.2)
3006921	Synaptophysin	110 (2.4)	32 (1.5)
3032734	MUM-1	48 (1.0)	55 (2.6)

Database and Relationships Following the OMOP CDM Format

The CDM's NOTE_NLP table is composed of three types of pathology reports; the major entities extracted from the immunohistochemical study reports and molecular study reports contain result summaries in the CDM's MEASUREMENT table. Surgical site and cancer diagnosis information extracted from the surgical specimen pathology reports was constructed by populating the cancer diagnosis information in the patient's CONDITION_OCCURRENCE table.

The FACT_RELATIONSHIP table in the OMOP CDM contains records about the relationships between facts stored as records in any table of the CDM. Relationships can be defined between facts from the same domain or different domains. Examples of fact relationships include personal relationships (parent-child), care site relationships (hierarchical organizational structure of facilities within a health system), indication relationships (between drug exposures and associated conditions), usage relationships (of devices during the course of an associated procedure), or facts derived from one another (measurements derived from an associated specimen). To define the relationship between the MEASUREMENT, CONDITION_OCCURRENCE, and SPECIMEN tables, the FACT_RELATIONSHIP table of the CDM was populated (Figure 5).

Figure 5. Overall data table relationships developed by the study.



NOTE_NLP: Pathology reports of surgical specimen

COLUMN_NAME	VALUE
NOTE_NLP_ID	30003371
NOTE_ID	200814772
SECTION_CONCEPT_ID	3025891 (Pathology report final diagnosis Narrative)
SNIPPET	Adenocarcinoma (by who 2010)
OFFSET	959
LEXICAL_VARIANT	Adenocarcinoma, nos
NOTE_NLP_CONCEPT_ID	0
NOTE_NLP_SOURCE_CONCEPT_ID	44498738 (Adenocarcinoma, nos)
NLP_SYSTEM	SNUBH_HIRC_NLPv1
NLP_DATE	2018-07-17
TERM_EXISTS	TRUE
TERM_MODIFIERS	0

COLUMN_NAME	VALUE
NOTE_NLP_ID	30003372
NOTE_ID	200814772
SECTION_CONCEPT_ID	3008885 (Pathology report site of origin Narrative)
SNIPPET	Rectum and sigmoid colon, anterior resection
OFFSET	908
LEXICAL_VARIANT	Sigmoid colon
NOTE_NLP_CONCEPT_ID	0
NOTE_NLP_SOURCE_CONCEPT_ID	44498127 (Sigmoid colon; C18.7)
NLP_SYSTEM	SNUBH_HIRC_NLPv1
NLP_DATE	2018-07-17
TERM_EXISTS	TRUE
TERM_MODIFIERS	0

NOTE_NLP: Molecular study report

COLUMN_NAME	VALUE
NOTE_NLP_ID	2004100
NOTE_ID	201558134
SECTION_CONCEPT_ID	3001274 (NRAS gene mutations found [Identifier] in Blood or Tissue by Molecular genetics method Nominal)
SNIPPET	Codon 61 (exon 3) | Negative
OFFSET	733, 741
LEXICAL_VARIANT	Codon 61 (exon 3)
NOTE_NLP_CONCEPT_ID	0
NOTE_NLP_SOURCE_CONCEPT_ID	0
NLP_SYSTEM	SNUBH_HIRC_NLPv1
NLP_DATE	2018-07-17
TERM_EXISTS	FALSE
TERM_MODIFIERS	negated=TRUE; value_as_concept_id=9189; value_as_narrative=Negative)

NOTE_NLP: Immunohistochemical study report

COLUMN_NAME	VALUE
NOTE_NLP_ID	1004897
NOTE_ID	200773014
SECTION_CONCEPT_ID	40758358 (Immune stain study)
SNIPPET	EGFR (GI) | 1+/3
OFFSET	400, 410
LEXICAL_VARIANT	EGFR
NOTE_NLP_CONCEPT_ID	3016231 (Epidermal growth factor receptor Ag [Presence] in Tissue by Immune stain)
NOTE_NLP_SOURCE_CONCEPT_ID	0
NLP_SYSTEM	SNUBH_HIRC_NLPv1
NLP_DATE	2018-07-17
TERM_EXISTS	FALSE
TERM_MODIFIERS	negated=FALSE; value_as_concept_id=9189; value_as_narrative=1+/3

CONDITION_OCCURRENCE

COLUMN_NAME	VALUE
CONDITION_OCCURRENCCE_ID	1610001772
PERSON_ID	257534973675956
CONDITION_CONCEPT_ID	4200514 (Adenocarcinoma of sigmoid colon)
CONDITION_START_DATE	2018-07-17
CONDITION_END_DATE	2018-07-17
CONDITION_TYPE_CONCEPT_ID	32535 (Tumor Registry)
CONDITION_SOURCE_VALUE	8140/3-C18.7 (Adenocarcinoma, NOS, of sigmoid colon)
CONDITION_SOURCE_CONCEPT_ID	44504380 (Adenocarcinoma, NOS, of sigmoid colon)

MEASUREMENT

COLUMN_NAME	VALUE
MEASUREMENT_ID	3400001148
PERSON_ID	257534973675956
MEASUREMENT_CONCEPT_ID	3001274 (NRAS gene mutations found [Identifier] in Blood or Tissue by Molecular genetics method Nominal)
MEASUREMENT_DATE	2018-07-17
MEASUREMENT_TYPE_CONCEPT_ID	32423 (NLP derived)
VALUE_AS_CONCEPT_ID	9189 (Negative)
MEASUREMENT_SOURCE_VALUE	N-RAS gene mutation test
VALUE_SOURCE_VALUE	Codon 12 (exon 2) = Negative; Codon 13 (exon 2) = Negative; Codon 61 (exon 3) = Negative;
MODIFIER_OF_EVENT_ID	1610001772 (CONDITION_OCCURRENCE_ID)
MODIFIER_FIELD_CONCEPT_ID	1147127 (condition_occurrence.condition_occurrence_id)

MEASUREMENT

COLUMN_NAME	VALUE
MEASUREMENT_ID	340007358
PERSON_ID	257534973675956
MEASUREMENT_CONCEPT_ID	3016231 (Epidermal growth factor receptor Ag [Presence] in Tissue by Immune stain)
MEASUREMENT_DATE	2018-07-17
MEASUREMENT_TYPE_CONCEPT_ID	32423 (NLP derived)
VALUE_AS_CONCEPT_ID	9191 (Positive)
MEASUREMENT_SOURCE_VALUE	EGFR
VALUE_SOURCE_VALUE	1+/3
MODIFIER_OF_EVENT_ID	1610001772 (CONDITION_OCCURRENCE_ID)
MODIFIER_FIELD_CONCEPT_ID	1147127 (condition_occurrence.condition_occurrence_id)

FACT_RELATIONSHIP

COLUMN_NAME	VALUE
DOMAIN_CONCEPT_ID_1	19 (Condition)
FACT_ID_1	1610001772
DOMAIN_CONCEPT_ID_2	36 (Specimen)
FACT_ID_2	1099738697
RELATIONSHIP_CONCEPT_ID	4038483 (Specimen source identity (SNOMED))

COLUMN_NAME	VALUE
DOMAIN_CONCEPT_ID_1	36 (Specimen)
FACT_ID_1	1099738697
DOMAIN_CONCEPT_ID_2	19 (Condition)
FACT_ID_2	1610001772
RELATIONSHIP_CONCEPT_ID	4038483 (Specimen source identity (SNOMED))

SPECIMEN

COLUMN_NAME	VALUE
SPECIMEN_ID	1099738697
PERSON_ID	257534973675956
SPECIMEN_CONCEPT_ID	0
SPECIMEN_TYPE_CONCEPT_ID	581378 (EHR Detail)
SPECIMEN_DATE	2018-07-17
SPECIMEN_SOURCE_ID	180717588901

FACT_RELATIONSHIP

COLUMN_NAME	VALUE
DOMAIN_CONCEPT_ID_1	26 (Measurement)
FACT_ID_1	3400001148
DOMAIN_CONCEPT_ID_2	36 (Specimen)
FACT_ID_2	1099738697
RELATIONSHIP_CONCEPT_ID	4023266 (Has specimen (SNOMED))

COLUMN_NAME	VALUE
DOMAIN_CONCEPT_ID_1	36 (Specimen)
FACT_ID_1	1099738697
DOMAIN_CONCEPT_ID_2	26 (Measurement)
FACT_ID_2	3400001148
RELATIONSHIP_CONCEPT_ID	4038483 (Specimen source identity (SNOMED))

In Figure 5, in the patient's surgical specimen pathology reports, records regarding the surgical site (the sigmoid colon) and diagnosis ("adenocarcinoma, nos") are stored. These entities are delivered into the CONDITION_OCCURRENCE table, with cancer diagnosis information (adenocarcinoma of the sigmoid colon). Specimen information, including person_id and specimen_id during surgery, is stored in the SPECIMEN table. Here, the relationship definition is made using a FACT_RELATIONSHIP for the specimen with a diagnosis of CONDITION_OCCURRENCE. We used the "specimen source identity" concept to define and link the relationship between the two tables. The results are also stored in NOTE_NLP while performing the molecular study and immunohistochemical study tests with the specimen obtained from the operation. In our example, *NRAS* mutation test results in the molecular study report were negative; its details are "Codon 61 (exon 3)". In addition, the information "NLP_derived N-RAS gene mutations found [identifier] in blood or tissue by molecular genetics method nominal" was updated in the MEASUREMENT table. In the same way, the immunohistochemical study test result for the EGFR test ("Epidermal growth factor receptor Ag [Presence] in Tissue by immune stain"), including the contents of the result "1+/3", was saved in the NOTE_NLP table first; then, the NLP_derived EGFR positive summary is delivered to the MEASUREMENT table.

Discussion

Principal Findings

We demonstrate that we can store colon cancer–related textual entities processed through NLP, including those from concepts of standard vocabularies (eg, SNOMED CT, LOINC, and ICD-O), in CDM tables that can later be used for computational phenotyping. Our system can thus enable the development of new standard feature-based NLP systems and the reuse, portability, adaptation, and extension of other cancer-based reports. This is the first study to provide details on how to extract major entities and standardize each item in pathology examination reports. However, variations in text data used at different institutions complicate the application of our pattern-matching rule to other observational health care databases.

Concept recognition in biomedical text has been addressed by multiple systems such as MetaMap [17], the ConceptMapper Annotator of the Unstructured Information Management Architecture (UIMA) [18], and the Clinical Text Analysis and Knowledge Extraction System (cTAKES) [19]. MetaMap is a widely available program that provides access to concepts in the UMLS metathesaurus from biomedical text; it was introduced in an effort to improve biomedical text retrieval, specifically the retrieval of MEDLINE/PubMed citations. It also links the text of biomedical literature to knowledge, including synonymy relationships, embedded in the metathesaurus [17]. ConceptMapper Annotator was implemented as a UIMA component. It was designed to accurately map text onto controlled vocabularies, specified as dictionaries, including the association of any necessary properties from the controlled vocabulary as part of that mapping [18]. Although these tools allow the use of a custom dictionary, their approach to efficient task performance relies on the target text being grammatically well formed. Thus, these mappers rely on token-based matching, which is processed by sentence and phrase segmentation. Many of the abbreviations, typing errors, and implicit tables in clinical text blur sentence boundaries and make phrase segmentation challenging. Our proposed approach utilizes rule-based pattern matching and a set of terminology mappings among international standard vocabularies such as LOINC, SNOMED CT, and ICD-O that link both concepts and values.

Many previous studies have focused on developing new text processing technologies. For instance, Baghari et al [20] developed a UMLS-based biomedical semantic operator. Solt [21] developed a medication extraction system using combined conditional random fields and rule-based systems. NOBLE Coder [22] implements a general algorithm for matching terms to concepts from an arbitrary vocabulary set. Its developers benchmarked the system’s speed and accuracy against the Colorado Richly Annotated Full Text (CRAFT) [23,24] and Shared Annotated Resources (ShARe) [25] corpora as reference standards and compared it to other concept recognition systems for biomedical tasks. Moreover, the BioCreAtIvE (Critical Assessment of Information Extraction in Biology) competition for automated gene and protein name recognition consists of a community-wide effort to evaluate information extraction and text mining developments in the biological domain [26]. Mitsumori et al used the support vector machine algorithm as a learning method for gene and protein name recognition [27], investigating and evaluating the system’s performance when making partial dictionary pattern matches.

Regarding research on text engineering and CDMs, there has been a recent report of a cohort retrieval system that can execute textual cohort selection queries on both structured and unstructured EHR data using CDM data [28]. The system leveraged a combination of structured queries and information retrieval techniques in NLP results to improve cohort retrieval performance while adopting OMOP CDM to enhance model portability. The NLP component empowered by cTAKES was used to extract CDM concepts from textual queries, and a hierarchical index in Elasticsearch was generated to support CDM concept search using information retrieval techniques and frameworks [28].

Our study approach differs from those in previous studies based on the following characteristics. First, we analyzed three kinds of pathology reports and created pattern-matching rules for each. Second, we converted the colorectal cancer entities of colon cancer patients in our hospital into international standards to establish them in the OMOP CDM database. Third, a mapping dictionary was created to standardize terms for pathology reports; this is a reusable legacy asset and can be used in future studies. To our knowledge, this is the first time that relationships have been defined using FACT_RELATIONSHIP to link OMOP CDM to other tables; these are also recyclable assets that can be implemented in other studies.

In our study, text data extraction errors can occur during CDM conversion from the pathology reports, and errors may also be introduced in standard term mapping. To prevent pathology text data extraction errors, we had to check the correctness of all items to be extracted. Text extraction results of the NOTE_NLP table and the number of data stored in the CDM database table were compared to confirm errors in the CDM table. From randomly sampled documents, we measured the accuracy of each item extracted by comparing the original document with the NLP extraction result and confirmed that there were no errors. Standard term mapping errors were checked for by domain experts. Mapping errors were identified through expert reviews and double-checking by CDM researchers.

From this study, we derived the insight that domain experts such as pathologists and bioinformatics specialists are essential for accurate term mapping, as understanding genetic testing methods is required to examine pathology reports. Furthermore, it was difficult to automate mapping tasks because some terms could be mapped to many terms (1:N); another difficulty was confusion from incorrect standard term search results. Another important insight of this study is that the NOTE_NLP table does not have clear standard guidelines for modifiers, so further standardization studies on term modifiers are needed to express various kinds of information in the future.

There were several challenges faced in this study. More general application of the pattern-matching rules created in this study is difficult because report formats vary for every institution. In our hospital's surgical pathology report, the results of examination of the pathological tissues of various organs removed during surgery are written in one report. In this study, one surgical site and diagnostic information were extracted from one surgical pathology report, except for multi-organ results. This will have to be supplemented in further research. The text extraction process deals with the contents of one hospital's examination report and corresponding rules, which can make this system difficult to use in other studies; however, it is expected that the definition of the extracted entities and the standard term mapping data table will be available for other studies as well. Although text extraction is not universal, the mapping guidelines can be adjusted for reusability because it is possible to modify them to apply to other organizations. Since it is important to convert the data in the same standard concept_id in order to conduct multi-institution studies using CDMs, we expect that even small mapping sets that include the histologic type of invasive carcinoma, location information, and biomarker term information will be easier to map into standard terminology. In addition, this will be more useful because the data are from reports generated at a tertiary general hospital. It will be fully available as a reference for other institutions. Additionally, the textual components of cancer pathology in clinical documents are as diverse as the various noncolon cancer types, and they vary by location and over time with new forms and templates; therefore, further research is necessary for the normalization of different types of cancer records and for the establishment of an extended database. In addition, there are several ways to interpret the results of the immunohistochemical study report for colorectal cancer. In our study, biomarker results such as EGFR were mostly found in the results of the immunohistochemical study report of colorectal cancer. The EGFR expression and interpretation could be diverse based on pathologists' visual scoring or the fraction of carcinoma staining. In particular, the pathology department of our hospital chose the scoring method that maps EGFR results of 1+, 2+, and 3+ as “positive,” so we converted accordingly. If the scoring method should be subdivided into different biomarker results such as C-erbB2, this is a limitation of our research and something we should do in a future study.

In this study, we extracted fundamental text entities from pathological examination reports of patients with colon cancer and built a CDM database through terminology standardization and database definition. The preparation of pathology data for cancer and genome research, as well as various textual data that are currently recorded and managed, could lead to various elemental problems; further research on the CDM is expected to utilize much data. Our system can extract and store key features from unstructured text as NLP annotations by using the format defined by the OMOP CDM. The essential text entities from pathology reports were extracted, standardized, and deployed. Furthermore, more sophisticated preparation of the pathology data is needed for further research on cancer genomics, and various types of text narratives are areas for additional research on the use of data in the CDM.

Limitations

This study could not be validated for an external organization for CDM. However, we attempted to build a consolidated CDM for dealing with narrative text, which enables researchers to more easily derive important information from unstructured data. Furthermore, with our concept-relation definition approach, other organizations can construct database structures. We hope to derive and apply more features of cancer-related text entities to utilize both clinical and omics data in further clinical studies based on this study.

Acknowledgments

This work was supported by the Bio & Medical Technology Development Program of the National Research Foundation of Korea funded by Korea’s Ministry of Science and ICT (NRF2017M3A9B6062319). This work was also supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program) (20003883, Advancing and expanding CDM-based distributed biohealth data platform) funded by Korea’s Ministry of Trade, Industry & Energy.

Authors' Contributions

BR analyzed the data and drafted the manuscript as the first author. EY contributed to the development of the NLP. SK, HB, and S Yi helped in preparing and evaluating the data. HN, SL, JWK, and HH contributed to pathology data discussions. RB initiated this study as the principal investigator of this project. S Yoo supervised the overall study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pseudo code for processing colorectal cancer pathology report.
[DOCX File , 73 KB - jmir_v22i12e18526_app1.docx]

Multimedia Appendix 2
Supplementary material for Immunochemistry concept mapping table.
[DOCX File , 16 KB - jmir_v22i12e18526_app2.docx]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Kim B, Kim JM, Kang GH, Chang HJ, Kang DW, Kim JH, et al. Standardized Pathology Report for Colorectal Cancer, 2nd Edition. J Pathol Transl Med 2020 Jan;54(1):1-19 [FREE Full text] [doi: 10.4132/jptm.2019.09.28] [Medline: 31722452]
3. Standards for Cancer Registries Volume V: Pathology Laboratory Electronic Reporting. URL: https://www.naaccr.org/wp-content/uploads/2016/11/Pathology-Laboratory-Electronic-Reporting-Version-4.0-April-2011.pdf [accessed 2019-06-12]
4. Sharma H, Mao C, Zhang Y, Vatani H, Yao L, Zhong Y, et al. Developing a portable natural language processing based phenotyping system. BMC Med Inform Decis Mak 2019 Apr 04;19(Suppl 3):78 [FREE Full text] [doi: 10.1186/s12911-019-0786-z] [Medline: 30943974]
5. Crepaz R, Romeo C, Montanaro D, De Santis S. Long-term results of treatment with bosentan in adult Eisenmenger's syndrome patients with Down's syndrome related to congenital heart disease. BMC Cardiovasc Disord 2013 Sep 18;13:74 [FREE Full text] [doi: 10.1186/1471-2261-13-74] [Medline: 24047157]

6. Brown SH, Lincoln MJ, Groen PJ, Kolodner RM. VistA--U.S. Department of Veterans Affairs national-scale HIS. Int J Med Inform 2003 Mar;69(2-3):135-156. [doi: 10.1016/s1386-5056(02)00131-4] [Medline: 12810119]
7. Brown SH, Lincoln M, Hardenbrook S, Petukhova ON, Rosenbloom ST, Carpenter P, et al. Derivation and evaluation of a document-naming nomenclature. J Am Med Inform Assoc 2001;8(4):379-390 [FREE Full text] [doi: 10.1136/jamia.2001.0080379] [Medline: 11418545]
8. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013 Oct;15(10):761-771 [FREE Full text] [doi: 10.1038/gim.2013.72] [Medline: 23743551]
9. Rasmussen LV, Kiefer RC, Mo H, Speltz P, Thompson WK, Jiang G, et al. A Modular Architecture for Electronic Health Record-Driven Phenotyping. AMIA Jt Summits Transl Sci Proc 2015;2015:147-151 [FREE Full text] [Medline: 26306258]
10. FitzHenry F, Resnic F, Robbins S, Denton J, Nookala L, Meeker D, et al. Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership. Appl Clin Inform 2015;6(3):536-547 [FREE Full text] [doi: 10.4338/ACI-2014-12-CR-0121] [Medline: 26448797]
11. Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. J Biomed Inform 2016 Dec;64:333-341 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.016] [Medline: 27989817]
12. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc 2012;19(1):54-60 [FREE Full text] [doi: 10.1136/amiajnl-2011-000376] [Medline: 22037893]
13. Park RW. Sharing Clinical Big Data While Protecting Confidentiality and Security: Observational Health Data Sciences and Informatics. Healthc Inform Res 2017 Jan;23(1):1-3 [FREE Full text] [doi: 10.4258/hir.2017.23.1.1] [Medline: 28261525]
14. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]
15. Goldsmith J, Siegal G, Suster S, Wheeler T, Brown R. Reporting guidelines for clinical laboratory reports in surgical pathology. Arch Pathol Lab Med 2008 Oct;132(10):1608-1616 [FREE Full text] [doi: 10.1043/1543-2165(2008)132[1608:RGFCLR]2.0.CO;2] [Medline: 18834219]
16. Patterson OV, Freiberg MS, Skanderson M, J Fodeh S, Brandt CA, DuVall SL. Unlocking echocardiogram measurements for heart disease research through natural language processing. BMC Cardiovasc Disord 2017 Jun 12;17(1):151 [FREE Full text] [doi: 10.1186/s12872-017-0580-8] [Medline: 28606104]
17. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
18. Tanenblatt M, Coden A, Sominsky I. The ConceptMapper Approach to Named Entity Recognition Internet. 2010 Presented at: Seventh International Conference on Language Resources and Evaluation (LREC'10); May 19-21, 2010; Valletta, Malta URL: http://www.lrec-conf.org/proceedings/lrec2010/pdf/448_Paper.pdf
19. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
20. Cuzzola J, Jovanović J, Bagheri E. RysannMD: A biomedical semantic annotator balancing speed and accuracy. J Biomed Inform 2017 Jul;71:91-109 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.016] [Medline: 28552401]
21. Tikk D, Solt I. Improving textual medication extraction using combined conditional random fields and rule-based systems. J Am Med Inform Assoc 2010;17(5):540-544 [FREE Full text] [doi: 10.1136/jamia.2010.004119] [Medline: 20819860]
22. Tseytlin E, Mitchell K, Legowski E, Corrigan J, Chavan G, Jacobson RS. NOBLE - Flexible concept recognition for large-scale biomedical natural language processing. BMC Bioinformatics 2016 Jan 14;17:32 [FREE Full text] [doi: 10.1186/s12859-015-0871-y] [Medline: 26763894]
23. Bada M, Eckert M, Evans D, Garcia K, Shipley K, Sitnikov D, et al. Concept annotation in the CRAFT corpus. BMC Bioinformatics 2012 Jul 09;13:161 [FREE Full text] [doi: 10.1186/1471-2105-13-161] [Medline: 22776079]
24. CRAFT: The Colorado Richly Annotated Full Text Corpus. URL: http://bionlp-corpora.sourceforge.net/CRAFT/ [accessed 2019-06-13]
25. ShARe/CLEF eHealth Evaluation Lab. URL: https://sites.google.com/site/shareclefehealth/home [accessed 2019-06-13]
26. Hirschman L, Yeh A, Blaschke C, Valencia A. Overview of BioCreAtIvE: critical assessment of information extraction for biology. BMC Bioinformatics 2005;6 Suppl 1:S1 [FREE Full text] [doi: 10.1186/1471-2105-6-S1-S1] [Medline: 15960821]
27. Mitsumori T, Fation S, Murata M, Doi K, Doi H. Gene/protein name recognition based on support vector machine using dictionary as features. BMC Bioinformatics 2005;6 Suppl 1:S8 [FREE Full text] [doi: 10.1186/1471-2105-6-S1-S8] [Medline: 15960842]
28. Liu S, Wang Y, Wen A, Wang L, Hong N, Shen F, et al. Implementation of a Cohort Retrieval System for Clinical Data Repositories Using the Observational Medical Outcomes Partnership Common Data Model: Proof-of-Concept System Validation. JMIR Med Inform 2020 Oct 06;8(10):e17376 [FREE Full text] [doi: 10.2196/17376] [Medline: 33021486]

Abbreviations

CDM: common data model
cTAKES: Clinical Text Analysis and Knowledge Extraction System
EGFR: epidermal growth factor receptor
EHR: electronic health record
ICD-O: International Classification of Diseases for Oncology
LOINC: Logical Observation Identifiers Names and Codes
NLP: natural language processing
NOS: not otherwise specified
OHDSI: Observational Health Data Sciences and Informatics
OMOP: Observational Medical Outcomes Partnership
SNOMED: Systematized Nomenclature of Medicine
SNOMED CT: SNOMED Clinical Terms
SNUBH: Seoul National University Bundang Hospital
UIMA: Unstructured Information Management Architecture
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 04.03.20; peer-reviewed by RW Park, Y Li; comments to author 25.03.20; revised version received 20.05.20; accepted 11.11.20; published 09.12.20.

Please cite as:
Ryu B, Yoon E, Kim S, Lee S, Baek H, Yi S, Na HY, Kim JW, Baek RM, Hwang H, Yoo S
Transformation of Pathology Reports Into the Common Data Model With Oncology Module: Use Case for Colon Cancer
J Med Internet Res 2020;22(12):e18526
URL: https://www.jmir.org/2020/12/e18526
doi:10.2196/18526
PMID:33295294

Original Paper

Comparing Precision Machine Learning With Consumer, Quality, and Volume Metrics for Ranking Orthopedic Surgery Hospitals: Retrospective Study

Dev Goyal[1], MSc; John Guttag[2], PhD; Zeeshan Syed[1], PhD; Rudra Mehta[1], BSc; Zahoor Elahi[1], MSc; Mohammed Saeed[1,3], MD, PhD

[1]Health at Scale Corporation, San Jose, CA, United States

[2]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, United States

[3]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Mohammed Saeed, MD, PhD
Department of Internal Medicine
University of Michigan
1500 East Medical Center Drive
Ann Arbor, MI, 48109
United States
Phone: 1 6176421280
Email: msaeed@umich.edu

Abstract

Background: Patients' choices of providers when undergoing elective surgeries significantly impact both perioperative outcomes and costs. There exist a variety of approaches that are available to patients for evaluating between different hospital choices.

Objective: This paper aims to compare differences in outcomes and costs between hospitals ranked using popular internet-based consumer ratings, quality stars, reputation rankings, average volumes, average outcomes, and precision machine learning–based rankings for hospital settings performing hip replacements in a large metropolitan area.

Methods: Retrospective data from 4192 hip replacement surgeries among Medicare beneficiaries in 2018 in a the Chicago metropolitan area were analyzed for variations in outcomes (90-day postprocedure hospitalizations and emergency department visits) and costs (90-day total cost of care) between hospitals ranked through multiple approaches: internet-based consumer ratings, quality stars, reputation rankings, average yearly surgical volume, average outcome rates, and machine learning–based rankings. The average rates of outcomes and costs were compared between the patients who underwent surgery at a hospital using each ranking approach in unadjusted and propensity-based adjusted comparisons.

Results: Only a minority of patients (1159/4192, 27.6% to 2078/4192, 49.6%) were found to be matched to higher-ranked hospitals for each of the different approaches. Of the approaches considered, hip replacements at hospitals that were more highly ranked by consumer ratings, quality stars, and machine learning were all consistently associated with improvements in outcomes and costs in both adjusted and unadjusted analyses. The improvement was greatest across all metrics and analyses for machine learning–based rankings.

Conclusions: There may be a substantive opportunity to increase the number of patients matched to appropriate hospitals across a broad variety of ranking approaches. Elective hip replacement surgeries performed at hospitals where patients were matched based on patient-specific machine learning were associated with better outcomes and lower total costs of care.

(J Med Internet Res 2020;22(12):e22765) doi:10.2196/22765

KEYWORDS

machine learning; hospital ratings; precision delivery; hospital; surgery; outcome; perioperative; internet; reputation; machine learning

Introduction

Patients undergoing elective surgeries often seek information at different levels of granularity when choosing providers, ranging from institutions to practices to individual physicians. It is widely established that the choice of provider significantly impacts both perioperative outcomes and costs [1-4]. There exist a variety of approaches that are available to patients for evaluating between different provider choices. These include consumer ratings, government quality ratings (eg, Centers for Medicare & Medicaid Services [CMS] stars), reputation rankings, and average volumes and outcomes [5-17]. Machine learning–based online tools are emerging as an alternative method of predicting provider performance that can factor patient-specific characteristics into provider rankings [17,18]. Unfortunately, there is little prior research into the comparative performance of these methods in predicting outcomes associated with different providers [19-21].

This study compares different approaches (consumer ratings, CMS quality stars, reputation rankings, and average volumes and outcomes) and personalized patient-specific rankings using machine learning (referred to as precision navigation) for predicting hospital performance, as measured by postprocedure hospitalization rate, emergency department (ED) visits, and total cost of care. The analyses were performed on data from Medicare fee-for-service (FFS) beneficiaries following elective hip replacement surgery. For each of the different approaches, 90-day outcomes for patients that were treated at highly ranked institutions were measured and compared to the outcomes for patients treated at institutions that were not highly ranked. Outcomes following elective hip replacement surgeries were studied because they are among the most commonly performed surgeries and because there is considerable variation in risk-adjusted performance across hospitals. Postprocedure hospitalization, ED visits, and total cost of care were used as outcome measures because they are quantifiable, meaningful to patients, and available for Medicare FFS populations.

Methods

Study Population

The study population comprised Medicare FFS beneficiaries undergoing elective hip replacements between 2013 to 2018 in the Chicago-Naperville-Arlington Heights, Illinois metropolitan area. Hip replacement surgeries were identified among these patients using standard procedure codes available in administrative claims data.

All hospitals in the greater Chicago metropolitan area that were visited for hip replacement surgery by at least one Medicare FFS beneficiary who lived in the Chicago-Naperville-Arlington Heights, Illinois area between 2016 and 2018 were included in the analysis (thus, for all hospitals that were active between 2016 and 2018, all data from between 2013 and 2018 were used in the analysis). For each patient, up to 10 of the nearest candidate hospitals that were a maximum of 50 miles from their place of surgery were ranked. For each patient, the top-ranked choice was determined for the different approaches, and the average outcomes and costs for patients who had surgeries at a top-ranked hospital were compared with the population averages.

Ranking Methodologies

Ratings

For each hospital, online consumer ratings from Yelp [6], Healthgrades [22], and US News [19] and quality star ratings from CMS Hospital Compare [10] were used for ranking purposes. The 2019 ratings available at the time of conducting this study were used for each approach in the absence of historical values for these approaches. Yelp and CMS Hospital Compare included overall hospital ratings scored between 1 and 5 (rather than specialty-specific ratings). Healthgrades online reports included ratings for hospital performance specific to hip replacement procedures. US News ratings for orthopedic surgery for each hospital were used. If two or more hospitals had the same ratings, the hospital with the larger number of reviews (or patient volume in the case of CMS ratings) was considered the top-ranked hospital.

Statistics

Procedural volumes were calculated for hip replacement using Medicare FFS data for each hospital between 2013 and 2017. The average postprocedure hospitalization rate and ED admission rate at each hospital following hip replacement procedures were calculated using Medicare FFS data as a measure of quality and were also treated as a rating. For each of these ratings (procedural volume and average rates), a single average was computed from 2013 to 2017. For evaluating volume-based ratings, the top hospital was the one with the highest average yearly volume of hip replacement surgeries. For evaluating average outcome–based ratings, the top hospital was the one with the lowest combined average 90-day postprocedure hospitalization rate and ED admission rate for hip replacement surgery.

Precision Navigation

Machine learning–based rankings of orthopedic facilities were generated for each beneficiary undergoing hip replacement in 2018 using a commercially available software system (Precision Navigation; Health at Scale Corp) that uses precision navigation and was trained on data prior to 2018 (ie, from 2013-2017) [18]. The top hospital for each beneficiary was determined as the top-ranked result returned by this system based on the patient's individual characteristics. The system predicts the hospitals likely to have the best long-term outcomes for patients based on the patient's personalized medical characteristics by learning from historical data about hospitals' long-term outcomes on similar patients.

Outcomes

The 90-day postprocedure hospitalization rate and rate of ED admission following the surgery were measured as outcomes. In addition, the 90-day total cost of care (TCOC) was also measured for each patient, which included the costs of the surgery and subsequent costs for the patient that occurred over the next 90 days from the day of surgery. The TCOC estimate included reimbursements for inpatient, outpatient, skilled nursing facility, and home health agency care, as these were available for the FFS beneficiaries. All patients included in the analyses had at least 90 days of follow-up (ie, follow-up data for patients were available until April 2019).

The observed average outcomes (postprocedure hospitalization rate, ED admission) and TCOC were compared between patients who received treatment at the top-ranked hospital for hip replacement surgery in 2018 and the overall population (all patients who had hip replacement surgery in 2018) in an unadjusted analysis. The unadjusted comparisons were also done in subgroups stratified by the Elixhauser comorbidity score. In each comparison, we statistically compared the outcomes of interest between patients who were admitted to a top-ranked hospital and patients who were admitted to a lower-ranked hospital (as determined by each respective method).

Pairwise comparisons between patients who visited a top-ranked hospital based on the different approaches were made using an adjusted propensity-based analysis with a one-to-one matching caliper without replacement [23]. Propensity was calculated using the weighted Elixhauser comorbidity score (each Elixhauser comorbidity was weighted to calculate a composite score for hospital readmissions) [24]. The propensity-matched groups for each pairwise comparison were generated 100 times by sampling with replacement. Thus, within each pairwise comparison between the various rating approaches, only patients with similar comorbidity scores (based on the matching caliper) were compared.

Results

Table 1 presents the characteristics of the Medicare FFS beneficiaries in the Chicago metropolitan area that received an inpatient hip replacement in 2018. The data consisted of 4192 hip replacement surgeries performed at 69 total hospitals.

Table 2 presents descriptive summaries for consumer ratings, quality stars, reputation rankings, average yearly surgical volume, average outcome rates, and precision navigation–based ranking for hospitals in the Chicago metropolitan area. The majority of the approaches had ratings for all of the hospitals in the region (all approaches excluding US News and Healthgrades, which covered more than half but not all of the hospitals in the Chicago metropolitan area).

Table 3 compares the outcomes and TCOC of patients who had surgeries at top hospitals with those of the overall population for each approach. Patients who underwent surgeries at hospitals that were ranked as top hospitals according to the CMS quality stars, Yelp, the average volume, the average outcome rate, and the precision navigation–based approaches were associated with substantially better outcomes than the population averages. This improvement in outcomes was greatest for precision navigation. The US News and Healthgrades approaches were associated with improvements in ED admission rates and postprocedure hospital rates or in TCOC but not in both. The precision navigation–based rankings were associated with an improvement of 3.4% for postprocedure hospitalization rate ($P<.001$), 4.1% for ED admission rates ($P<.001$), and US $3315 for TCOC ($P<.001$) between top hospitals and the overall population. The other ranking methods were associated with smaller variations in outcomes between top hospitals and the overall population. The percentage of patients who underwent surgery at a hospital ranked as a top hospital by each ratings approach varied from 27.6% (1159/4192) for the precision navigation–based rankings to 49.6% (2078/4192) for the average volume–based rankings.

Table 1. Patient characteristics used in the hip replacement analyses.[a]

Patient characteristics	Value
Patients, n	4192
Age (years), mean (SD)	69.4 (8.2)
Male, n (%)	1589 (38)
Race/ethnicity, n (%)	
White	3602 (86)
African American	417 (10)
Unknown	85 (2)
Other	39 (1)
Hispanic	28 (1)
Asian	21 (1)
Comorbidities, n (%)[b]	
Hypertension	1699 (43)
Anemic deficiency	428 (10)
Diabetes	653 (16)
Hypothyroidism	398 (9)
Chronic lung disease	374 (9)
Obesity	320 (8)
Electrolyte imbalance	284 (7)
Depression	273 (7)
Arthritis	235 (6)
Tumors without metastases	223 (5)
90-day outcomes[c]	
ED[d] admission rate, n (%)	655 (15.6)
Postprocedure hospitalization rate, n (%)	469 (11.2)
Reimbursement (US $), mean (SD)	27,000 (21,000)

[a] The study cohort includes all Medicare fee-for-service beneficiaries in the Chicago metropolitan area that underwent an elective hip replacement surgery in 2018.

[b] The 10 most common comorbidities are shown.

[c] ED and postprocedure hospitalizations are calculated as occurring at least once within the 90-day follow-up period.

[d] ED: emergency department.

Table 2. A summary of the ranking methods used in this study.

Rating system and characteristic	Value
US News	
Hospitals, n	60
Range	0-100
Mean (SD)	42 (10)
Yelp	
Hospitals, n	87
Range	1-5
Number of reviews, mean	57
Mean rating (SD)	2.5 (0.7)
Healthgrades	
Hospitals, n	75
Range	1-5
Mean rating (SD)	2.5 (1.2)
CMS[a]	
Hospitals, n	90
Range	1-5
Mean (SD)	3.1 (1.3)
Average volumes	
Hospitals, n	94
Range	0-752
Mean rating (SD)	153 (156)
Average outcomes: ED[b] **admission rate**	
Hospitals, n	94
Range	5-100
Mean rating (SD)	18.2 (11.0)
Average outcomes: postprocedure hospitalization rate	
Hospitals, n	94
Range	0-50
Mean hip rating (SD)	14.7 (8.1)
Precision navigation	
Hospitals, n	90
Range	1-10

[a] CMS: Centers for Medicare & Medicaid Services.

[b] ED: emergency department.

Table 3. Comparison of average outcomes between patients who went to top-ranked hospitals for hip replacement surgery and the overall population.

Outcome	Precision navigation	US News	Healthgrades	Yelp	CMS[a]	Average volume	Average outcome rate
90-day ED[b] admission rate							
In-top average[c], n (%)	486.3 (11.6)	639.8 (15.3)	659.0 (15.7)	585.3 (14.0)	584.4 (13.9)	605.2 (14.4)	584.6 (13.9)
Population average[d] – in-top average, n (%)	171.9 (4.1)	15.2 (0.4)	–4.0 (0.1)	69.7 (1.7)	70.6 (1.7)	49.8 (1.2)	70.4 (1.7)
P value	<.001	.55	.91	.03	.009	.04	.03
90-day postprocedure hospitalization rate							
In-top average, n (%)	327.0 (7.8)	413.0 (9.9)	472.6 (11.3)	407.4 (9.7)	396.7 (9.5)	417.6 (10.0)	415.6 (9.9)
Population average – in-top average, n (%)	142.5 (3.4)	56.0 (1.3)	–3.6 (0.1)	61.6 (1.5)	72.3 (1.7)	51.4 (1.2)	53.4 (1.3)
P value	<.001	.01	.91	.03	.002	.01	.05
90-day total cost of care							
In-top average (US $)	23,698	28,328	26,944	28,088	26,412	28,047	28,067
Population average – in-top average (US $)	3315	–1315	68	1076	601	1035	1054
P value	<.001	<.001	.005	<.001	.003	<.001	<.001
Patients who visited top hospitals, n (%)	1159 (27.6)	1959 (46.7)	1304 (31.1)	1461 (34.9)	1786 (42.6)	2078 (49.6)	1513 (36.1)

[a] CMS: Centers for Medicare & Medicaid Services.

[b] ED: emergency department.

[c] In-top average: outcome rate among patients who were admitted to top-ranked hospitals using the particular ranking approach.

[d] Population average: average value of the particular outcome in the overall population.

Table 4 presents the average outcomes and TCOC for patients who had surgeries at top hospitals, stratified by their Elixhauser comorbidity scores. For both low (≤0) and high (>0) comorbidity score groups, patients who visited top hospitals using precision navigation–based ranking had the best outcomes and TCOC.

Table 5 compares the outcomes and TCOC of patients in a pairwise, adjusted, propensity-matched analysis. For each pairwise comparison, patients matched to top hospitals using both ranking approaches were propensity matched, and the differences in average outcome rates between these propensity-matched patients were computed. In the pairwise propensity-matched comparison using the precision navigation–based approach with each of the other approaches, the improvement in outcomes was greatest for precision navigation for every outcome and every pairwise comparison.

Table 4. Comparison of outcomes between patients who went to top-ranked hospitals for hip replacement surgery and the general population, using different ranking methodologies.

Elixhauser comorbidity score quantile and outcome	Precision navigation	US News	Healthgrades	Yelp	CMS[a]	Average volume	Average outcome rate
≤0 (low)							
Patients who visited top hospitals, n (%)	933.7 (29.4)	1461.0 (46.0)	990.9 (31.2)	1133.8 (35.7)	1349.8 (42.5)	1559.4 (49.1)	1140.2 (35.9)
90-day ED[b] admission rate							
In-top average[c], n (%)	327.1 (10.3)	390.6 (12.3)	463.7 (14.6)	371.6 (11.7)	374.8 (11.8)	368.4 (11.6)	362.1 (11.4)
In-bin average – in-top average, n (%)	98.5 (3.1)	31.8 (1.0)	–41.3 (1.3)	54.0 (1.7)	47.6 (1.5)	54.0 (1.7)	60.3 (1.9)
P value	<.001	.13	.16	.04	.03	.005	.02
90-day postprocedure hospitalization rate							
In-top average, n (%)	187.4 (5.9)	219.1 (6.9)	289.0 (9.1)	222.3 (7.0)	225.5 (7.1)	209.6 (6.6)	241.4 (7.6)
In-bin average – in-top average, n (%)	88.9 (2.8)	54.0 (1.7)	–12.7 (0.4)	54.0 (1.7)	50.8 (1.6)	63.5 (2.0)	34.9 (1.1)
P value	<.001	.002	.59	.01	.007	<.001	.10
90-day TCOC[d]							
In-top average (US $)	22,963	26,696	26,237	27,045	25,032	26,404	26,713
In-bin average – in-top average (US $)	2665	–1068	–610	–1418	595	–776	–1086
P value	<.001	<.001	.13	<.001	.001	<.001	<.001
>0 (high)							
Patients who visited top hospitals, n (%)	224.5 (22.1)	498.9 (49.1)	311.9 (30.7)	328.2 (32.3)	434.8 (42.8)	520.2 (51.2)	373.9 (36.8)
90-day ED admission rate							
In-top average, n (%)	536.7 (16.9)	755.9 (23.8)	609.8 (19.2)	698.7 (22.0)	651.1 (20.5)	727.3 (22.9)	689.2 (21.7)
In-bin average – in-top average, n (%)	184.2 (5.8)	–34.9 (1.1)	111.2 (3.5)	25.4 (0.8)	73.0 (2.3)	–3.2 (0.1)	34.9 (1.1)
P value	.02	.41	.09	.75	.15	.94	.59
90-day postprocedure hospitalization rate							
In-top average, n (%)	495.5 (15.6)	584.4 (18.4)	581.2 (18.3)	609.8 (19.2)	533.6 (16.8)	635.2 (20.0)	543.1 (17.1)
In-bin average – in-top average, n (%)	111.2 (3.5)	22.2 (0.7)	25.4 (0.8)	–3.2 (0.1)	73.0 (2.3)	–28.6 (0.9)	63.5 (2.0)
P value	.15	.63	.73	.99	.11	.47	.25
90-day TCOC							
In-top average (US $)	26,748	33,102	29,192	31,691	30,696	32,969	32,189
In-bin average – In-top average (US $)	4593	–1761	2149	–350	645	–1628	–848
P value	<.001	<.001	.003	.63	.64	.003	<.001

[a] CMS: Centers for Medicare & Medicaid Services.

[b] ED: emergency department.

[c] In-top average: outcome rate among patients who were admitted to top-ranked hospitals using the particular approach.

[d] TCOC: total cost of care.

Table 5. A pairwise propensity-matched comparison of patients matched to top hospitals using each of the ranking approaches.

Reduction in outcome rate[a]	Relative to US News		Relative to Healthgrades		Relative to Yelp		Relative to CMS[b]		Relative to average volume		Relative to average outcome rate	
	Outcome	*P* value	Outcome	*P* value	Outcome	*P* value	Outcome	*P* value	Outcome	*P* value	Outcome	*P* value
Precision navigation												
90-day ED[c] admission rate, %	2.5	<.001	3.8	<.001	2.0	<.001	1.8	<.001	2.1	<.001	1.7	<.001
90-day postprocedure hospitalization rate, %	1.1	<.001	2.9	<.001	1.4	<.001	1.2	<.001	1.1	<.001	1.6	<.001
90-day TCOC[d] (US $)	4025	<.001	2923	<.001	4298	<.001	2309	<.001	3840	<.001	3817	<.001
US News												
90-day ED admission rate, %	N/A[e]	N/A	0.4	<.001	–0.9	<.001	–1.3	<.001	–0.7	<.001	–1.1	<.001
90-day postprocedure hospitalization rate, %	N/A	N/A	1.5	<.001	–0.1	.87	–0.3	.89	0.3	.43	0.0	.22
90-day TCOC (US $)	N/A	N/A	–1161	<.001	–269	.32	–1918	<.001	–191	<.001	–299	.001
Healthgrades												
90-day ED admission rate, %	0.4	<.001	N/A	N/A	–1.9	<.001	–1.7	<.001	–0.9	<.001	–1.7	<.001
90-day postprocedure hospitalization, %	1.5	<.001	N/A	N/A	–1.5	<.001	–1.8	<.001	–1.2	<.001	–1.5	<.001
90-day TCOC (US $)	–1161	<.001	N/A	N/A	757	<.001	–542	<.001	867	<.001	910	<.001
Yelp												
90-day ED admission rate, %	–0.9	<.001	–1.9	<.001	N/A	N/A	–0.2	.39	0.4	.002	–0.3	.02
90-day postprocedure hospitalization rate, %	–0.1	.87	–1.5	<.001	N/A	N/A	–0.5	<.001	0.1	.33	0.0	.30
90-day TCOC (US $)	–269	.32	757	<.001	N/A	N/A	–1693	<.001	–105	.19	–216	.16
CMS												
90-day ED admission rate, %	–1.3	<.001	–1.7	<.001	–0.2	.39	N/A	N/A	0.2	<.001	–0.1	.91
90-day postprocedure hospitalization rate, %	–0.3	.89	–1.8	<.001	–0.5	<.001	N/A	N/A	0.4	<.001	0.3	.004
90-day TCOC (US $)	–1918	<.001	–542	<.001	–1693	<.001	N/A	N/A	1486	<.001	1575	<.001
Average volume												
90-day ED admission rate, %	–0.7	<.001	–0.9	<.001	0.4	.002	0.2	<.001	N/A	N/A	–0.5	<.001
90-day postprocedure hospitalization rate, %	0.3	.43	–1.2	<.001	0.1	.33	0.4	<.001	N/A	N/A	–0.1	.05
90-day TCOC (US $)	–191	<.001	867	<.001	–105	.19	1486	<.001	N/A	N/A	–104	.86

[a] Differences were calculated as column – row; thus, positive values imply a lower adverse outcome rate.

[b] CMS: Centers for Medicare & Medicaid Services.

[c] ED: emergency department.

[d] TCOC: total cost of care.

[e] N/A: not applicable.

The potential impact of using different rating systems under the counterfactual assumption that patients only presented to a top hospital for them was also considered. The precision navigation–based ranking approach resulted in significantly more hospitals being designated as a top hospital for one or more Medicare FFS beneficiary (ie, distributed or balanced the case load across a larger number of hospitals that were top ranked for specific patients). Out of 69 total hospitals visited by patients in 2018 for hip replacements, 54 were considered top-ranked hospitals for at least one patient when using precision navigation. CMS ratings resulted in 25 hospitals being top ranked. US News and Healthgrades each resulted in 27 hospitals being top ranked. Yelp-based ranking resulted in 22 hospitals being considered top ranked. Average volume and outcome resulted in 26 and 24 hospitals, respectively, being considered top-ranked hospitals for a patient.

Discussion

Prior research on the use of popular ranking and rating approaches, including web-based ratings, consumer guides, and various quality ratings for physicians or hospitals, have resulted in inconsistent findings, and it is unclear which rating approach works best [5,20,21,25-28]. This retrospective study compares the performance of several different rating strategies for designating top hospitals for a large population of Medicare FFS beneficiaries who underwent elective hip replacement surgeries in the Chicago metropolitan area. The study also compared the performance of the aforementioned approaches for hospital rankings and ratings with a more personalized precision navigation–based approach that selects hospitals based on patients’ individual health characteristics.

Overall, several approaches were shown to be associated with better outcomes and lower TCOC when patients presented to a top hospital based on the respective ranking approach. These included CMS quality stars and precision navigation–based rankings, with top-ranked hospitals achieving improved outcomes and TCOC for hip replacement using both propensity-adjusted and nonadjusted analyses. The greatest improvements were observed for precision navigation–based rankings, which were more consistently associated with reductions in 90-day postprocedure hospitalization rate, ED admission rate, and TCOC in each analysis.

Prior research has resulted in an inconsistent correlation between outcomes and ratings in top-ranked hospitals. For example, Cram et al [29] showed no significant differences in total knee arthroplasty outcomes in top-ranked and non–top-ranked hospitals using the US News rankings. Studies in other surgical subspecialties, such as cardiac surgery, found only a weak correlation between online ratings and perioperative mortality [25]. Osborn et al [27] showed a significant correlation between favorable rankings and lower mortality rates associated with various major surgical procedures. However, use of mortality as a surrogate for outcome may not reflect other pertinent outcomes that occur with higher frequency, such as postprocedure hospital admissions or ED visits.

In this study, hospital performance was measured using the rates of postprocedure hospitalizations and ED visits and the total cost of care among Medicare FFS beneficiaries undergoing elective hip replacements. These metrics, which were available for all hospitals included in this study, are meaningful to patients. Many of the popular consumer-based ranking approaches are based more on patient satisfaction than on objective measures [6,25,26,30]. Velasco et al [31] found that negative online comments about orthopedic surgeons were associated with surgery-independent factors, such as waiting time and logistics. Austin et al [32] showed that there was considerable variation in the rankings of top hospitals when different criteria were used. The significant variation in ratings across the different rating platforms may complicate the choice of institution for patients [7,8,33]. Moreover, all of these approaches implicitly assume that the choice of a top hospital is independent of the characteristics of the patient [19,32]. The results of this study call this assumption into question.

Typical methods for ranking hospitals for specialty procedures assume that the quality of an institution is the same for all patients. The personalized approach of precision navigation–based ranking predicts outcomes for each hospital-patient pair by learning from the respective hospital’s prior outcomes for patients who resemble the patient being matched. Thus, this study suggests that hospital quality may be a personalized, variable phenomenon rather than a global, uniform value. The overwhelming majority of hospitals considered are not consistently ideal (ie, the top choice) for all patients but instead lie largely in a “gray zone” of being selectively good or bad for individual patients. More than 85% (80/94) of the hospitals that were included in the Chicago metropolitan area fell in this category. This finding is perhaps not surprising, given the complexity of surgical patients and the variations in case mixes, resources, training, and other organizational characteristics of hospitals that lead to facilities performing well or poorly for specific individuals. Prior research has shown that patient complexity is strongly correlated with outcomes in elective surgeries, such as total knee arthroplasty [34]. Bozic et al [35] also found considerable variation in respective patient population acuities of hospitals performing elective hip and knee replacements and a fourfold difference in risk-adjusted complication rates [35].

The potential policy implications of our study are significant. The paradigm of patients choosing institutions based on crowdsourced online ratings, popular consumer guides, and established metrics, such as high-volume centers, as surrogates for high performance may not be suitable for identifying the preferred centers for patients contemplating elective surgery.

Average outcome–based measures are more reliable than online consumer rankings but still far from optimal, and they did not result in lower TCOC. Our results suggest that a personalized approach based on precision navigation that uses readily available data to characterize a patient's medical complexity in the context of individual hospitals may be associated with substantial improvements in outcomes while also lowering TCOC. An additional policy concern is that any mechanism designed to encourage patients to go to specific hospitals should ideally balance hospital use and capacity. Prioritizing hospital choice based on the typical static rating approaches would lead to a greater concentration of surgeries in a small number of top hospitals, thus overwhelming certain hospitals while underutilizing others. The personalized approach distributed patients over more hospitals (54 hospitals) than any of the other approaches (27 or fewer hospitals). Finally, the low proportion of patients that were matched to top hospitals using the precision navigation–based approach (<30%) presents a substantial opportunity for improving outcomes and costs for many patients by steering them to hospitals that are best suited for them.

One of the limitations of this study is the incomplete characterization of each patient's unique medical needs and medical history, since only CMS administrative claims were available. Ideally, the online ratings would be based on reviews collected before the date of the procedure. However, access to historic ratings for online web-based systems was not available. Furthermore, the use of ED visits and postprocedure hospitalizations as outcomes does not reflect on the quality of life improvement following elective surgery that patients ultimately seek. This study was retrospective in nature and thus only showed associations between outcomes and various ranking and rating methods rather than suggesting causal relationships between ranking methods and patient outcomes. The ranking methods that were evaluated in this study's experimental design may not have been intended to be used as stand-alone referral services or offer advice on the suitability or choice of particular hospitals. Rather, they may have been intended to be used (where applicable) as additional information that can be factored in with other inputs (such as referral recommendations provided by patients' care providers) in choosing a hospital site for a future surgery. This study only evaluated the association of these ranking methodologies within the experimental framework considered, and the findings here should not be taken as a statement of their suitability for any purpose. Patients' choices in elective surgeries can be influenced by socioeconomic factors as well as factors beyond their control, such as the availability of surgeons, recommendations from their care providers, and the network limitations of their health insurance plans.

The present study was performed using Medicare beneficiary claims data that primarily comprised older patients, and their applicability to younger patients with private health care insurance or Medicaid requires further research. Another limitation of this work was that the analysis was restricted to the greater Chicago metropolitan area. Future work is needed to extend this study to other procedures, non-Medicare populations, and hospitals nationwide. A prospective trial is warranted to further study the impact of hospital-ranking approaches on patient outcomes and total cost of care.

Conflicts of Interest

Authors MS, ZS, and JG are cofounders, board members, and equity holders of Health at Scale, a machine learning software company. Health at Scale software was used for the precision machine learning algorithm that was evaluated in this study. DG, ZE, and RM are employees and hold equity in Health at Scale.

References

1. Khuri SF. The NSQIP: a new frontier in surgery. Surgery 2005 Nov;138(5):837-843. [doi: 10.1016/j.surg.2005.08.016] [Medline: 16291383]
2. Street A, Gutacker N, Bojke C, Devlin N, Daidone S. Variations in outcome and costs among NHS providers for common surgical procedures: econometric analyses of routinely collected data. Health Serv Deliv Res 2014;2(1):1-89. [doi: 10.3310/hsdr02010] [Medline: 25642520]
3. Miller DC, Gust C, Dimick JB, Birkmeyer N, Skinner J, Birkmeyer JD. Large variations in Medicare payments for surgery highlight savings potential from bundled payment programs. Health Aff (Millwood) 2011 Nov;30(11):2107-2115 [FREE Full text] [doi: 10.1377/hlthaff.2011.0783] [Medline: 22068403]
4. Gani F, Hundt J, Daniel M, Efron JE, Makary MA, Pawlik TM. Variations in hospitals costs for surgical procedures: inefficient care or sick patients? Am J Surg 2017 Jan;213(1):1-9. [doi: 10.1016/j.amjsurg.2016.05.007] [Medline: 27392753]
5. Lee V. Transparency and Trust - Online Patient Reviews of Physicians. N Engl J Med 2017 Jan 19;376(3):197-199. [doi: 10.1056/NEJMp1610136] [Medline: 28099823]
6. Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care. Health Affairs 2016 Apr;35(4):697-705. [doi: 10.1377/hlthaff.2015.1030]
7. Ellimoottil C, Kadlec AO, Farooq A, Quek ML. Choosing a physician in the Yelp era. Bull Am Coll Surg 2013 Nov;98(11):20-21. [Medline: 24313135]
8. Ellimoottil C, Leichtle SW, Wright CJ, Fakhro A, Arrington AK, Chirichella TJ, et al. Online physician reviews: the good, the bad and the ugly. Bull Am Coll Surg 2013 Sep;98(9):34-39. [Medline: 24455818]

9. Prabhu A, Randhawa S, Clump D, Heron D, Beriwal S. What Do Patients Think About Their Radiation Oncologists? An Assessment of Online Patient Reviews on Healthgrades. Cureus 2018 Feb 06;10(2):e2165 [FREE Full text] [doi: 10.7759/cureus.2165] [Medline: 29644154]
10. Bilimoria KY, Barnard C. The New CMS Hospital Quality Star Ratings: The Stars Are Not Aligned. JAMA 2016 Nov 01;316(17):1761-1762. [doi: 10.1001/jama.2016.13679] [Medline: 27802552]
11. Jha AK. The Stars of Hospital Care: Useful or a Distraction? JAMA 2016 Jun 07;315(21):2265-2266. [doi: 10.1001/jama.2016.5638] [Medline: 27272569]
12. Harder B, Comarow A. JAMA 2015 May 19;313(19):1903-1904. [doi: 10.1001/jama.2015.4566] [Medline: 25988456]
13. Finks JF, Osborne NH, Birkmeyer JD. Trends in Hospital Volume and Operative Mortality for High-Risk Surgery. N Engl J Med 2011 Jun 02;364(22):2128-2137. [doi: 10.1056/nejmsa1010705]
14. Sehgal AR. Ann Intern Med 2010 Apr 20;152(8):521-525. [doi: 10.7326/0003-4819-152-8-201004200-00009] [Medline: 20404383]
15. Toomey PG, Teta AF, Patel KD, Ross SB, Rosemurgy AS. High-volume surgeons vs high-volume hospitals: are best outcomes more due to who or where? Am J Surg 2016 Jan;211(1):59-63. [doi: 10.1016/j.amjsurg.2015.08.021] [Medline: 26542187]
16. Wray CM, Vali M, Walter LC, Lopez L, Austin PC, Byers A, et al. Examining the Utility of 30-day Readmission Rates and Hospital Profiling in the Veterans Health Administration. J Hosp Med 2019 May;14(5):266-271. [doi: 10.12788/jhm.3155] [Medline: 30794141]
17. The 'eharmony' of healthcare — how Health at Scale Technologies can match patients to the right providers for optimal outcomes. Becker's Hospital Review. 2019. URL: https://tinyurl.com/y3lcbe7k [accessed 2019-12-17]
18. Zhan T, Goyal D, Guttag J, Mehta R, Elahi Z, Syed Z, et al. Machine intelligence for early targeted precision management and response to outbreaks of respiratory infections. Am J Manag Care 2020 Oct;26(10):445-448 [FREE Full text] [doi: 10.37765/ajmc.2020.88456] [Medline: 33094940]
19. Gray BM, Vandergrift JL, Gao GG, McCullough JS, Lipner RS. Website ratings of physicians and their quality of care. JAMA Intern Med 2015 Feb;175(2):291-293. [doi: 10.1001/jamainternmed.2014.6291] [Medline: 25437252]
20. Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, et al. Web-Based Physician Ratings for California Physicians on Probation. J Med Internet Res 2017 Aug 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]
21. Daskivich TJ, Houman J, Fuller G, Black JT, Kim HL, Spiegel B. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. J Am Med Inform Assoc 2018 Apr 01;25(4):401-407. [doi: 10.1093/jamia/ocx083] [Medline: 29025145]
22. Healthgrades. URL: https://www.healthgrades.com/quality/ratings-awards/reports/americas-best-hospitals [accessed 2020-10-21]
23. Austin PC. A comparison of 12 algorithms for matching on the propensity score. Stat Med 2014 Mar 15;33(6):1057-1069 [FREE Full text] [doi: 10.1002/sim.6004] [Medline: 24123228]
24. Ondeck NT, Bovonratwet P, Ibe IK, Bohl DD, McLynn RP, Cui JJ, et al. Discriminative Ability for Adverse Outcomes After Surgical Management of Hip Fractures. J Orthop Trauma 2018;32(5):231-237. [doi: 10.1097/bot.0000000000001140]
25. Liu JJ, Matelski J, Cram P, Urbach DR, Bell CM. Association Between Online Physician Ratings and Cardiac Surgery Mortality. Circ Cardiovasc Qual Outcomes 2016 Nov;9(6):788-791. [doi: 10.1161/circoutcomes.116.003016]
26. Nwachukwu BU, Adjei J, Trehan SK, Chang B, Amoo-Achampong K, Nguyen JT, et al. Rating a Sports Medicine Surgeon's "Quality" in the Modern Era: an Analysis of Popular Physician Online Rating Websites. HSS J 2016 Oct;12(3):272-277 [FREE Full text] [doi: 10.1007/s11420-016-9520-x] [Medline: 27703422]
27. Osborne NH, Nicholas LH, Ghaferi AA, Upchurch GR, Dimick JB. Do popular media and internet-based hospital quality ratings identify hospitals with better cardiovascular surgery outcomes? J Am Coll Surg 2010 Jan;210(1):87-92. [doi: 10.1016/j.jamcollsurg.2009.09.038] [Medline: 20123337]
28. Trehan SK, Nguyen JT, Marx R, Cross MB, Pan TJ, Daluiski A, et al. Online Patient Ratings Are Not Correlated with Total Knee Replacement Surgeon-Specific Outcomes. HSS J 2018 Jul;14(2):177-180 [FREE Full text] [doi: 10.1007/s11420-017-9600-6] [Medline: 29983660]
29. Cram P, Cai X, Lu X, Vaughan-Sarrazin M, Miller B. Total knee arthroplasty outcomes in top-ranked and non-top-ranked orthopedic hospitals: an analysis of Medicare administrative data. Mayo Clin Proc 2012 Apr;87(4):341-348 [FREE Full text] [doi: 10.1016/j.mayocp.2011.11.017] [Medline: 22469347]
30. Cua S, Moffatt-Bruce S, White S. Reputation and the Best Hospital Rankings: What Does It Really Mean? Am J Med Qual 2017;32(6):632-637. [doi: 10.1177/1062860617691843] [Medline: 28693335]
31. Velasco BT, Chien B, Kwon JY, Miller CP. Online Ratings and Reviews of American Orthopaedic Foot and Ankle Surgeons. Foot Ankle Spec 2020 Feb;13(1):43-49. [doi: 10.1177/1938640019832363] [Medline: 30795702]
32. Austin JM, Jha AK, Romano PS, Singer SJ, Vogus TJ, Wachter RM, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. Health Aff (Millwood) 2015 Mar;34(3):423-430. [doi: 10.1377/hlthaff.2014.0201] [Medline: 25732492]

33. Downing NS, Cloninger A, Venkatesh AK, Hsieh A, Drye EE, Coifman RR, et al. Describing the performance of U.S. hospitals by applying big data analytics. PLoS One 2017;12(6):e0179603 [FREE Full text] [doi: 10.1371/journal.pone.0179603] [Medline: 28662045]
34. Anis HK, Sodhi N, Vakharia RM, Scuderi GR, Malkani AL, Roche MW, et al. Cost Analysis of Medicare Patients with Varying Complexities Who Underwent Total Knee Arthroplasty [in press]. J Knee Surg 2019 Aug 28:e. [doi: 10.1055/s-0039-1695716] [Medline: 31461755]
35. Bozic KJ, Grosso LM, Lin Z, Parzynski CS, Suter LG, Krumholz HM, et al. Variation in Hospital-Level Risk-Standardized Complication Rates Following Elective Primary Total Hip and Knee Arthroplasty. J Bone Joint Surg Am 2014;96(8):640-647. [doi: 10.2106/jbjs.l.01639]

Abbreviations

CMS: Centers for Medicare & Medicaid Services
ED: emergency department
FFS: fee-for-service
TCOC: total cost of care

Edited by G Eysenbach, R Kukafka; submitted 22.07.20; peer-reviewed by Z Hydari, C Zhou; comments to author 09.08.20; revised version received 19.08.20; accepted 14.10.20; published 01.12.20.

Please cite as:
Goyal D, Guttag J, Syed Z, Mehta R, Elahi Z, Saeed M
Comparing Precision Machine Learning With Consumer, Quality, and Volume Metrics for Ranking Orthopedic Surgery Hospitals: Retrospective Study
J Med Internet Res 2020;22(12):e22765
URL: https://www.jmir.org/2020/12/e22765
doi:10.2196/22765
PMID:33258459

Original Paper

The Influence of Doctors' Online Reputation on the Sharing of Outpatient Experiences: Empirical Study

Yang Wang[1], PhD; Hong Wu[1], PhD; Xueqin Lei[1], BA; Jingxuan Shen[1], BA; Zhanchun Feng[1], PhD

School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Corresponding Author:
Hong Wu, PhD
School of Medicine and Health Management
Tongji Medical College
Huazhong University of Science and Technology
13 Hangkong Road, Qiaokou District
Hubei Province
Wuhan, 430030
China
Phone: 86 13277942186
Email: wuhong634214924@163.com

Abstract

Background: The internet enables consumers to evaluate products before purchase based on feedback submitted by like-minded individuals. Displaying reviews allows customers to assess comparable experiences and encourages trust, increased sales, and brand positivity. Customers use reviews to inform decision making, whereas organizations use reviews to predict future sales. Prior studies have focused on manufactured products, with little attention being paid to health care services. In particular, whether patients prefer to use websites to discuss doctors' reputation has so far remained unanswered.

Objective: This study aims to investigate how patient propensity to post treatment experiences changes based on doctors' online reputation (medical quality and service attitude) in delivering outpatient care services. Further, this study examines the moderating effects of hospitals' (organizational) online reputation and disease severity.

Methods: Fractional logistic regression was conducted on data collected from 7183 active doctors in a Chinese online health community to obtain empirical results.

Results: Our findings show that patients prefer to share treatment experiences for doctors who have a higher medical quality and service attitude ($\beta_{\text{service attitude}}$=.233; $P<.001$ and $\beta_{\text{medical quality}}$=.052; $P<.001$) and who work in hospitals with a higher online reputation (β=.001; $P<.001$). Patients are more likely to share experiences of doctors who treat less severe diseases, as opposed to those treating severe diseases (β=−.004; P=.009). In addition, hospitals' online reputation positively (negatively) moderates the relationship between medical quality (service attitude) and patient propensity to post treatment experiences, whereas the moderating effects of disease severity on doctors' online reputation are negative.

Conclusions: Our research contributes to both theory and practice by extending the current understanding of the impact of individual reputation on consumer behavior. We investigate the moderating effects of organizational reputation and consumer characteristics in online health communities.

(J Med Internet Res 2020;22(12):e16691) doi:10.2196/16691

KEYWORDS

online health communities; individual reputation; doctor reputation; patient feedback; organizational reputation; disease severity

Introduction

Background

In seeking health care provision, patients often face uncertainty regarding the quality of doctors' services, lacking trustworthy channels for accessing information such as medical quality and bedside manner [1]. Medical quality has historically been associated with treatments received by hospitals at the organizational level; however, patients are increasingly seeking information relating to the quality of individual doctors, that is, at the doctor level. Information asymmetries between patients and doctors exist extensively, with patients now regularly interacting on social networking sites to inform their provision needs based on peer recommendations. With the growing popularity of web 2.0 technologies, online health care communities provide a useful channel for patients to obtain doctors' information. In China, more than 80% of patients search for health care information before visiting hospitals [2].

Internet-based media play an important role in providing prepurchase information and informing decisions. These burgeoning new media have been hailed as a democratizing force that enables consumers to discuss products and services online [3]. In online communities, consumers critically evaluate the quality of comparable products by analyzing brands, pricing, and retailer reputation (note: in this paper, we use treatment experiences, reviews, word-of-mouth, and feedback interchangeably, as well as patients, consumers, and service receivers), which act as signals of product quality [4,5]. Prior literature suggests that a higher reputation can also signal higher quality [6,7]. Numerous empirical studies have suggested that reputation is one of the predominant factors influencing consumer purchases and seller performance [8,9], consistently revealing that there is a close link between consumer reviews and future sales. Sellers' online reviews positively impact product demand [10], with reviews creating a bridge of communication between consumers and sellers, decreasing consumers' perceived risk, and boosting trust and cooperation on both sides [11].

Reviews, which are generally agreed to be more effective than traditional advertising [12], increasingly affect consumer behaviors [13,14]. They facilitate the prediction of future performance and, therefore, are required more than ever by sellers. Existing literature shows that consumers who are pleased or displeased with a product will make their opinions known to others [15]; the more satisfied or dissatisfied the consumer, the more likely they are to post feedback about their experiences [16]. However, much less is known in relation to health care products and services. To the best of our knowledge, only Wu and Lu [17] have studied the role of doctors' reputation and its influence on patient propensity to share reviews, identifying that doctors with a higher reputation receive a greater number of reviews. However, their study focused solely on the role of the individual doctor's reputation and failed to consider the organizational reputation and consumer characteristics, which are important factors that affect consumer behaviors. According to the theory of psychological choice [18], the effect of a signal (individual reputation) is moderated by environmental situations (organizational reputation) and contextual factors (such as consumer characteristics), with final responses dependent on their interaction effects. This study aims to fill this critical gap by studying the impact factors of patient propensity to post treatment experiences online. We seek to understand and address the following questions:

- *Question 1*: How does a doctor's reputation influence patient propensity to post treatment experiences online?
- *Question 2*: How does a hospital's reputation moderate the relationship between the doctor's reputation and patient propensity to post treatment experiences online?
- *Question 3*: How does disease severity moderate the relationship between the doctor's reputation and patient propensity to post treatment experiences online?

We argue that reputation, signaled by existing reviews, can predict future reviews. Data were collected from an online health community, which, in recent years, has helped patients find doctors, book outpatient care services, and search for medical information. Unlike extant literature on manufactured products, our study includes both medical quality and service attitude, which are important factors in the health care field, as part of the doctor's reputation in our model [19]. In recent years, patients have complained about doctors' bad attitudes. Thus, in addition to improving service quality, it is vital to mitigate conflicts between doctors and patients by enhancing service attitude [20]. Moreover, we examined the moderating effects of hospital (organizational level) reputation and disease severity (patient characteristics). patient propensity to post treatment experiences is the ratio of the increment of the treatment experience to the increment of outpatient care service demands over a time window of interest.

Online Health Communities

In recent years, online health care communities have been developed by patient organizations, medical service providers, and nonprofit organizations to make it easier for patients to find health-related information [21]. Such communities provide virtual forums for patients to obtain services and discuss treatment experiences. Researchers have started to investigate the benefits and user behaviors of such communities, from the perspective of doctors [19,22] and patients [23-25]. For doctors, Ni and Sun [26] studied the willingness of doctors to work on online platforms and associated benefits. For patients, Xiao et al [27] explored whether patients' information search behavior influenced their perceived health condition. With regard to the impact factors of whether or not to post health information online, people take privacy and information sensitivity into consideration [28].

In China, as a result of continued limitations in existing health care provisions, online health communities have been strongly adopted by citizens. China has the world's largest population and thus represents a huge resource-consumption country. China's large population generates a variety of unique health care needs and, therefore, exhibits unique behaviors within online health care communities. Health ultimately concerns everyone, and with the emergence of online health care communities, patients now have more channels to find doctor information, whereas doctors have more choices in the way they deliver medical treatment. On the basis of extant literature, we have found few studies that explore the effects of doctors' reputation on patient propensity to post treatment experiences and the moderating effects of hospitals' reputation and disease severity. Our research, therefore, aims to fill these gaps.

Theory of Psychological Choice

Hansen [18] presented an overview of psychologists' approaches to consumer choice and the processes employed in different scenarios. He determined that the consumer choice process is characterized by conflict, uncertainty, cognitive activity, and related psychological processes. Individuals' choices depend on their current situation, whereas the nature of the problem can be described as a comprehensive result of internal and external factors. Behavioral response is a result of the interaction between environmental factors and individual characteristics such as values, beliefs, intentions, and preferences. According to the theory of psychological choice [18], consumer behavior, dependent on circumstances, is decided by a series of factors, including internal and external factors. The effect of a signal is moderated by environmental situations and contextual factors, such as consumer characteristics, with final responses being decided by the interaction effects of these factors.

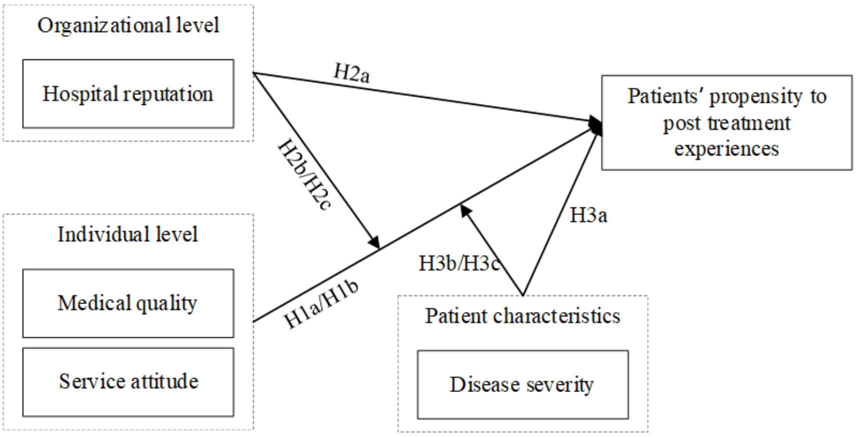
Expectation-Confirmation Theory

Expectation-confirmation theory is widely used to explore consumer behavior in both product marketing [29] and service marketing [30]. Oliver [31] described in detail the process of expectation, confirmation, and postpurchase behaviors. First, consumers form an initial expectation of the product or service, which is shaped by personal experiences, norms, and the present environment [32]. Second, consumers form perceptions about performance after receiving the product. Third, consumers will assess their pre-expectation and perceived performance and estimate the gap (degree of confirmation) between expectation and perceived performance. Fourth, the confirmation between pre-expectation and perceived performance influences their satisfaction and ultimately determines their future behaviors, such as repurchase intention and word-of-mouth.

Research Hypotheses

Consumer reviews are an important criterion that impacts consumer behavior. However, existing literature rarely investigates the relationship between doctors' online reputation and patient propensity to post treatment experiences online. We sought to examine how doctors' medical quality and service attitude affect their patient propensity to post treatment experiences. Moreover, we attempted to investigate the moderating effects of the hospital's reputation and disease severity. Figure 1 shows the conceptual model of this study.

Figure 1. Conceptual model.



Nowadays, the internet enables consumers to easily post opinions and express thoughts, feelings, and viewpoints on products and services to the wider online community [33]. On the basis of the expectation-confirmation theory, high reputation enhances consumers' expectations of quality and vice versa [34]. Pre-expectation will be compared with the perceived and/or actual performance received following the purchase of a product or service. Compared with low expectations, a high expectation is less likely to be reached by perceived performance [35]. The degree of confirmation between expectation and perceived performance dictates consumer satisfaction [31]. For consumers, the primary motivation to share positive or negative comments is to inform others [15] and/or to express their satisfaction or dissatisfaction [36]. Consumers are likely to express their feelings to others when their expectations are either surpassed or not met [37], and the propensity to post online reviews is greater for extreme experiences but smaller for average experiences [38].

In online health communities, patients hold comparatively high expectations about service quality for doctors with a high reputation. High expectations are less likely to be reached by perceived quality. The degree of expectation would affect consumer satisfaction and their propensity to post about treatment experiences. Higher expectations cause patients to be easily disappointed and dissatisfied after receiving services, which leads to them sharing negative feelings with others online [15,34]. The present literature has indicated that the existing reputation has a positive impact on the number of future reviews received [39-41]. However, the potential mechanism is unexamined in online health communities. On the basis of the abovementioned insights, we hypothesize the following:

- *Hypothesis 1a*: Doctors' medical quality positively affects patient propensity to post treatment experiences online.
- *Hypothesis 1b*: Doctors' service attitude positively affects patient propensity to post treatment experiences online.

An organization's reputation helps consumers make informed choices when they feel uncertain about a product or service [42]. Organizational reputation strongly influences consumer expectation and purchase intention [43]. In product marketing, Amblee and Bui [39] demonstrated that the amount of future online reviews has a positive correlation with a product's existing brand (organizational) reputation. In online health communities, a hospital's reputation can be considered a signal to patients. Patients would have higher expectations from doctors who work in hospitals with high online reputations. Higher expectations more easily induce disconfirmation between pre-expectation and perceived performance, which would enhance patient propensity to post treatment experiences.

- *Hypothesis 2a*: A hospital's reputation positively affects patient propensity to post treatment experiences online.

On the basis of the theory of psychological choice [18], the effect of a signal is influenced by environmental situations. As the medium for diffusing the signal, environments can influence the strength and effectiveness of the signal [44]. Many studies have examined the effect of signals in different environments and obtained consistent conclusions that the strength of a signal is moderated by signal environment uncertainty [45]. Therefore, the impact of individual reputation can be moderated by organizational reputation. A positive evaluation of an organization's reputation generates a positive evaluation of an individual's reputation [46].

With regard to online health communities, a hospital's reputation can be treated as an environmental factor. The delivery process of a signal varies among different hospital environments. Thus, the hospital's reputation can moderate the effect of a doctor's reputation. In reducing patients' perceived risks and increasing trust in the doctor's reputation, a higher hospital reputation can make patients have a higher expectation about doctors' performance. On the basis of the expectation-confirmation theory [34], higher expectation is less likely to be reached by perceived performance [35]. Patients will be more disappointed after receiving the service and are more likely to express their feelings to others online [15,34]. On the basis of the abovementioned insights, we develop the following hypotheses:

- *Hypothesis 2b*: A hospital's online reputation positively moderates the relationship between a doctor's medical quality and patient propensity to post treatment experiences online.
- *Hypothesis 2c*: A hospital's online reputation positively moderates the relationship between a doctor's service attitude and patient propensity to post treatment experiences online.

On the basis of the theory of psychological choice [18], the effect of a signal is also influenced by contextual factors. The influence of reputation varies with different types of products and services [47]. Individual characteristics significantly affect the degree of satisfaction with service quality [48] and moderate the relationship between service quality and satisfaction [49].

In the health care field, patient behavior is also influenced by their characteristics. Disease severity is an important basis for distinguishing between patients. Prior research has indicated that disease severity moderates the doctor's reputation on the patient's purchasing behavior [19]. It is argued that patients with severe diseases prefer to choose doctors with high medical quality rather than service attitude [19]. The study mainly focuses on the consumer buying process, but the decision on posting feedback is neglected.

From a positive perspective, disease severity may influence the patient's physical and mental health [50]. Patients with severe diseases are more sensitive to the doctor's reputation than those with less severe diseases. For example, compared with patients with less severe diseases, such as the common cold, patients with tumors are likelier to choose a doctor with a higher reputation and form higher expectations regarding their desire for health. Moreover, patients with severe diseases experience more pain and distress and are eager to find higher quality services.

From a negative perspective, patients with severe diseases often concentrate less on service attitude [51]. They prefer to choose doctors with higher medical quality rather than service attitude [19]. Purchasing preference leads to different feedback behaviors. Service attitude has fewer effects on patient propensity to post treatment experiences for patients with severe diseases. However, as a highly professional service, medical service sets an invisible barrier for patients who generally lack professional medical knowledge to assess medical quality, especially for severe diseases. Patients with severe diseases may not evaluate perceived quality objectively, thereby leading to a lack of certainty of disconfirmation between pre-expectation and perceived quality. In fact, given the issues they focus on, patients with severe diseases are more likely to have concerns about their recovery time and health, rather than posting feedback or complaining about poor experiences online. Furthermore, privacy concerns and information sensitivity, 2 critical influencing factors for deciding whether or not to spread health information online [28], are concerned by patients with severe diseases, ultimately decreasing patient propensity to post treatment experiences online.

On the basis of the aforementioned insights, we plan to determine the advantages of these effects in specific contexts. We propose both positive and negative moderating effects of disease severity:

- *Hypothesis 3a*: Disease severity significantly affects patient propensity to post treatment experiences online.
- *Hypothesis 3b*: Disease severity significantly moderates the relationship between a doctor's medical quality and patient propensity to post treatment experiences online.
- *Hypothesis 3c*: Disease severity significantly moderates the relationship between a doctor's service attitude and patient propensity to post treatment experiences online.

Methods

In this section, we describe the research context and data collection process and present the variables and models.

Research Context

We test our hypotheses using data collected from the WeDoctor website, a leading online health community authorized by the Chinese Health and Family Planning Committee. WeDoctor has become the leading online health community in China, mainly providing appointment booking services for outpatient care. The website helps increase efficiency for both patients and hospitals. Using the WeDoctor website, patients can easily make appointments and save valuable time. By 2020, the community has helped more than 850 million citizens. The WeDoctor website started to provide online written consultation and video consultation services in September 2016. In our proposed model, we do not include written and video consultation services for 2 reasons. First, compared with the outpatient care service appointment function, written and video consultation services are rarely used by patients. Second, our data were collected in the first half of 2016 when only outpatient care appointment services were provided by the website.

More than 7800 hospitals and 260,000 doctors are active in the online community. WeDoctor creates home pages for doctors and their hospitals. Doctors can self-manage their home pages, including modifying schedules for outpatient care services and updating individual information. The website has a formal and comprehensive reputation mechanism, which is important for this study. Patients can post their treatment experiences after receiving outpatient care services in the hospitals. Treatment experiences help future patients make better choices.

Sample and Data Collection

We used a crawler to automatically download doctors' information from the WeDoctor website using the following selection criteria. First, we crawled all active doctors who usually add or modify their outpatient care service information or other individual information (active doctors are recognized by WeDoctor). Second, we selected doctors who treat severe diseases and who treat relatively less severe diseases. Severe diseases include malignant tumors and heart and cerebrovascular diseases. Less severe diseases include endocrine, digestive, and nervous system diseases. The reasons for choosing these disease categories will be explained in detail in the following section. We repeated the collection process in 2 time periods: one week in March 2016 and another week in June 2016. We included in our analyses the doctors who were seen at both collection times, yielding a sample of 7183 doctors. For each doctor, we collected their reviews, reputation, and other relevant information (eg, hospital information). We also collected information on the medical departments with which the doctors were affiliated.

From each doctor's home page, we collected information posted about patients' experiences. Each patient can give a score to the doctor's medical quality and service attitude observed during treatment. Other patients can then read these reviews to make informed decisions. Figure 2 shows an example of a doctor's home page, whereas Figure 3 shows an example of a hospital's home page.

Figure 2. A doctor's home page on the WeDoctor website.

治疗效果： 93%满意

医生态度： 99%满意

[redacted]

职称：主任医师 / 教授

医院：复旦大学附属华山医院

科室：胰腺外科

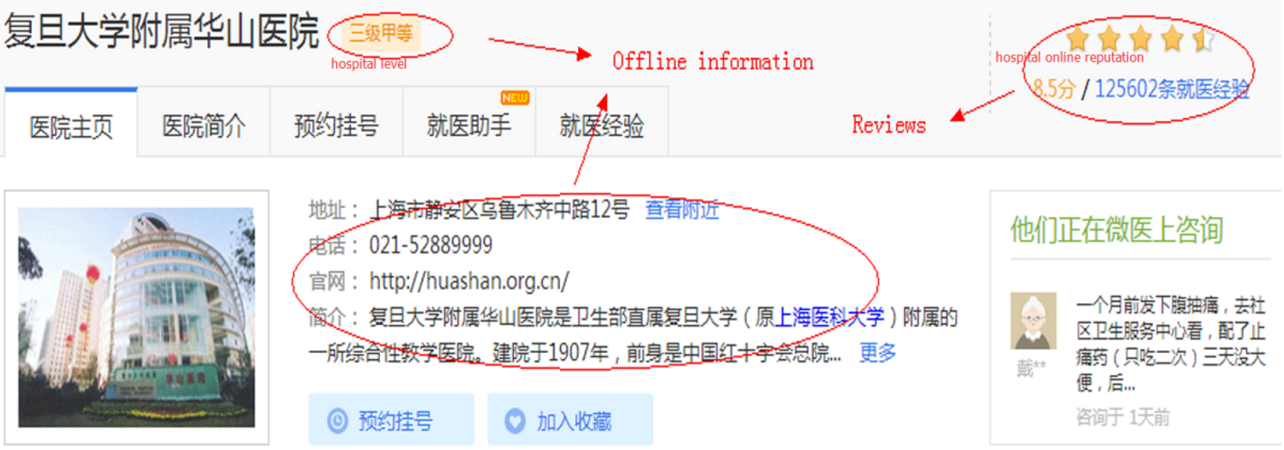
就诊原因：胰腺癌

疗效：满意 态度：很满意

预约后等待半小时，医生水平很高态度也好，不过得了癌症也得看个人的造化了！

王** 于 2014-09-16 13:53:12 (就诊后13天) 分享的就医经验

Figure 3. A hospital's home page on the WeDoctor website.



复旦大学附属华山医院 三级甲等

地址：上海市静安区乌鲁木齐中路12号 查看附近

电话：021-52889999

官网：http://huashan.org.cn/

简介：复旦大学附属华山医院是卫生部直属复旦大学（原上海医科大学）附属的一所综合性教学医院。建院于1907年，前身是中国红十字会总院... 更多

8.5分 / 125602条就医经验

他们正在微医上咨询

一个月前发下腹抽痛，去社区卫生服务中心看，配了止痛药（只吃二次）三天没大便，后...

戴** 咨询于 1天前

Variables and Empirical Models

The variables used in this study are in the form of aggregated data at the doctor level, which can help control for the potential influence of patient heterogeneity [52]. The detailed definitions for all variables included in this study are shown in Table 1.

Table 1. Variable definitions.

Variable	Definition
Dependent variables	
Patient propensity to post treatment experiences	The ratio of the increment of the treatment experience to the increment of outpatient care service demands over 3 months for each doctor.
Independent variables	
Medical quality	Patients give an evaluation score for doctors' medical quality when patients share treatment experiences. The WeDoctor calculates the mean of medical quality for each doctor based on all the existing treatment experiences posted by patients. The range of values for medical quality is from 0 to 1, with a greater value indicating a higher medical quality.
Service attitude	Patients give an evaluation score for doctors' service attitude when patients share treatment experiences. The WeDoctor calculates the mean of service attitude for each doctor based on all the existing treatment experiences posted by patients. The range of values for service attitude is from 0 to 1, with a greater value indicating a higher service attitude.
Moderating variables	
Hreputation	When patients post experiences, they also give a score on the hospital's environment and attitude of guide service. The range of values for the hospital's online reputation is from 0 to 10, with a greater value representing a higher level of satisfaction.
Disease_severity	The severity of disease that patients get. We use one dummy variable to measure it. When the disease is high-risk, the variable is equal to 1.
Control variables	
Dtitle_dummy1 and Dtitle_dummy2	Doctors' medical skills as evaluated by the government, including Chief Doctor, Associate Chief Doctor, and Attending Doctor. Two dummy variables were used. (0, 0) represents Attending Doctor title or below.
Hlevel_dummy	The variable indicating the comprehensive health care quality of doctor *i*'s affiliated hospital in terms of medical skills, equipment, human resources, etc. *Hlevel_dummy* presents AAA level and above hospitals. (0, 0) represents AA level hospital or below.

Dependent Variable

The dependent variable in our model is patient propensity to post treatment experiences. The variable is the ratio of the increment of the treatment experience to the increment of outpatient care service demands over a certain time period. The dependent variable is defined as follows:

where *i* represents each doctor. The subscripts *t* and *t−1* denote 2 periods in time and PPPTE denotes patient propensity to post treatment experiences.

Independent and Moderating Variables

The independent variable in our model is the doctor's online reputation, which is divided into 2 dimensions: medical quality and service attitude. The WeDoctor website calculates the mean of medical quality and mean of service attitude for each doctor based on all the existing treatment experiences posted by patients. The range of values for both medical quality and service attitude is from 0 to 1, with a greater value indicating a higher satisfaction.

The moderating variables were the hospital's online reputation and disease severity for patients treated in the hospital. The hospital's online reputation reflects the integral medical quality and integral service attitude delivered by the hospital. The range of values for the hospital's online reputation is from 0 to 10, with a greater value representing a higher level of satisfaction. We used mortality rates to distinguish the severity of different diseases. The Chinese Health Statistics Yearbook, published in 2019 [53], reports the latest health statistics, which lists mortality rates for different categories of diseases. We chose the first 3 fatal categories of diseases as severe diseases in our model. They are malignant tumor–related diseases (mortality rate: 163.18/100,000), heart diseases (mortality rate: 146.34/100,000), and cerebrovascular diseases (mortality rate: 128.88/100,000). For less severe diseases, we chose endocrine-related diseases (mortality rate: 21.15/100,000), digestive system diseases (mortality rate: 14.54/100,000), and nervous system diseases (mortality rate: 8.62/100,000). The difference in mortality rates between severe diseases and less severe diseases is large (nearly 10 times), which is helpful for understanding the impact of disease severity on patient behaviors. We collected all active doctors who treat these diseases, and finally, 5602 doctors who treat severe diseases and 1581 doctors who treat less severe diseases are included. A dummy variable was used to measure disease severity, with 1 representing severe diseases and 0 representing less severe diseases:

Control Variables

We included both doctors' titles and hospital levels in our model to control for their popularity offline. In China, each doctor has an offline title that represents their medical skills and level of experience, including Chief Doctor, Associate Chief Doctor, and Attending Doctor. These titles are evaluated and issued by government agencies. We use 2 dummy variables, Dtitle_dummy1 and Dtitle_dummy2, to measure doctors' titles. Similarly, each hospital in China is assigned a rank, classified as class A, B, or C, with class A being the best quality of hospital. Hospital level, which is also evaluated and issued by government agencies, represents their medical quality and medical technical strength. As the number of class C hospitals in this online health community is very small, we combined it with class B and used 1 dummy variable, Hlevel_dummy, to measure hospital level. The detailed definitions of these dummy variables are as follows:

We use general linear model regression to obtain empirical results. Fractional logistic regression is most suitable for our dependent variable (% of patients posting treatment experiences). On the basis of all the hypotheses, the empirical models are as follows:

$$Logit(PPPTE\_i)=\beta_{1,i}\ Dtitle\_dummy1+\beta_{2,i}\ Dtitle\_dummy2+\beta_{3,i}\ Hlevel\_dummy$$

$$+\beta_{4,i}\ Dmedical\_quality_{t-1}+\beta_{5,i}\ Dservice\_attitude_{t-1}+\beta_{6,i}\ Hreputation_{t-1}$$

$$+\beta_{7,i}\ Severity\_diseases+\beta_{8,i}\ Dmedical\_quality_{t-1}*Hreputation_{t-1}$$

$$+\beta_{9,i}\ Dservice\_attitude_{t-1}*Hreputation_{t-1}$$

$$+\beta_{10,i}\ Dmedical\_quality_{t-1}*Severity\_diseases$$

$$+\beta_{11,i}\ Dservice\_attitude_{t-1}*Severity\_diseases+\varepsilon$$

where *i* represents each doctor. The subscripts *t* and *t−1* denote 2 periods in time. We use data collected at time=*t−1* for the independent variable and time=*t* for dependent variables.

Results

Descriptive Statistics and Correlations

We use the expectation-confirmation theory in our hypotheses to argue that patients have higher expectations when they choose doctors with high reputations. Patients are likely to feel disconfirmed between expectation and perceived quality of the service and express their feelings online.

Table 2 presents descriptive statistics and correlations for the key variables in our study. We can see that doctors' reputation, hospitals' reputation, and disease severity are correlated with patient propensity to post treatment experiences. Doctors' reputation and the hospital's reputation positively affect patient propensity to post treatment experiences; conversely, disease severity negatively impacts patient propensity to post treatment experiences. All variance inflation factor values, of all variables, are below 5, which indicates the negligible effect of multicollinearity.

Table 2. Description and correlation (N=7183).

Variables	Mean (SD)	Minimum	Maximum	Patient propensity to post treatment experiences	Dtitle_dummy1	Dtitle_dummy2	Hlevel_dummy	Dmedical_quality	Dservice_attitude	Hreputation
Patient propensity to post treatment experiences	0.067 (0.108)	0	0.944	—[a]	—	—	—	—	—	—
Dtitle_dummy1	0.348 (0.476)	0	1	0.229[b]	—	—	—	—	—	—
Dtitle_dummy2	0.443 (0.496)	0	1	−0.050[b]	−0.591[b]	—	—	—	—	—
Hlevel_dummy	0.088 (0.283)	0	1	−0.180[b]	−0.016[b]	0.037[b]	—	—	—	—
Dmedical_quality	0.338 (0.442)	0	1	0.790[b]	0.197[b]	−0.038[b]	−0.167[b]	—	—	—
Dservice_attitude	0.402 (0.471)	0	1	0.860[b]	0.208[b]	−0.044[b]	−0.175[b]	0.889[b]	—	—
Hreputation	5.740 (4.192)	0	10	0.477[b]	0.082[b]	−0.055[b]	−0.328[b]	0.447[b]	0.486[b]	—
Severity_diseases	0.780 (0.414)	0	1	−0.329[b]	0.080[b]	−0.022[b]	−0.587[b]	0.280[b]	0.309[b]	0.537[b]

[a] This table is symmetrical. The number in the lower left corner is same as the at top right corner.

[b] Correlation is significant at the .01 level (2-tailed), significant at .01

Empirical Results

The analyses are deemed fit using Stata, a data analysis software. The empirical results are shown in Table 3. Model 1 contains all the control variables. Model 2 adds all the independent variables. Model 3 adds all the moderating variables, and model 4 adds all the interaction terms of the independent variables and moderating variables. We also tested the interactions between doctors' reputation, hospital reputation, and disease severity in model 5. As none of the interaction terms are significant, model 5 is not further discussed.

Hypotheses 1a and 1b concern the impact of the doctor's reputation on patient propensity to post treatment experiences. From model 4 in Table 3, we see that both medical quality and service quality have a positive impact on this dependent outcome. Patients are more likely to post reviews about their treatment experiences for those doctors who have a higher reputation. Moreover, the effect size of service attitude is nearly 4 times that of medical quality ($\beta_{\text{service attitude}}$=.233; *P*<.001 and $\beta_{\text{medical quality}}$=.052; *P*<.001), which indicates that service attitude plays a more important role in influencing patient propensity to post treatment experiences than medical quality; thus, both hypotheses 1a and 1b are supported.

As the results of model 4 show that a hospital's reputation has no significant impact, we focused on its margin effect, with results demonstrating that a hospital's reputation has a positive influence on patient propensity to post treatment experiences (β=.001; *P*<.001). Patients are more likely to post about their treatment experiences for those doctors who work in hospitals with a high reputation. Thus, hypothesis 2a is supported. Disease severity had a negative impact on patient propensity to post treatment experiences (β=.004; *P*=.009). People with more severe diseases are less likely to post reviews online. Thus, hypothesis 3a is supported.

Hypotheses 2b and 2c test the moderating effects of organizational reputation on the relationship between individual reputation and consumer behavior. From model 4 in Table 3, we observe that a hospital's reputation positively moderates the relationship between medical quality and patient propensity to post treatment experiences (β=.002; *P*=.01) and negatively moderates the relationship between service attitude and patient propensity to post treatment experiences (β=.004; *P*=.01). The impact of medical quality on patient propensity to post treatment experiences is greater for doctors who work in hospitals with higher reputations, whereas the impact of service attitude on patient propensity to post treatment experiences is smaller for doctors who work in hospitals with higher reputations; thus, hypothesis 2b is supported, whereas hypothesis 2c is not.

Hypotheses 3b and 3c examine the moderating effects of consumer characteristics (disease severity) on the relationship between individual reputation and consumer behavior. From model 4 in Table 3, we see that disease severity not only negatively moderates the relationship between medical quality (β=.036; *P*<.001) and patient propensity to post treatment experiences but also negatively moderates the relationship between service attitude (β=.044; *P*<.001) and patient propensity to post treatment experiences. The impact of both medical quality and service attitude on patient propensity to post treatment experiences is smaller for doctors who treat severe diseases; thus, both hypotheses 3b and 3c are supported.

To better interpret our results, we use the empirical results for the dependent variable, the increment of outpatient care service demands, and take its log value in the empirical model. The results are shown in Table 4. The impact of medical quality on patients' choice was greater for patients with severe diseases than for those with less severe diseases (β=.319; *P*=.04). Severe diseases increase the demand for outpatient care services and positively moderate the relationship between doctors' reputation and outpatient care service demands. On the contrary, for patient propensity to post treatment experiences, our results show that doctors who treat severe diseases are less likely to receive reviews about treatment experience (β=.325; *P*=.05). Moreover, the moderating effects of disease severity on doctors' reputation are negative.

Table 3. Results for the patient propensity to post treatment experiences. General linear model regression was used to obtain results.

Variables	Model 1, coefficient (SD)	Model 2, coefficient (SD)	Model 3, coefficient (SD)	Model 4, coefficient (SD)	Model 5, coefficient (SD)
Constant	0.054[a] (0.006)	0.012[b] (0.003)	0.026[a] (0.003)	0.012[a] (0.004)	0.012[a] (0.004)
Dtitle_dummy1	0.031[a] (0.015)	−0.012[a] (0.008)	−0.011[a] (0.008)	−0.010[a] (0.008)	−0.010[a] (0.008)
Dtitle_dummy2	0.015[a] (0.015)	−0.006[a] (0.008)	−0.005[a] (0.008)	−0.004[a] (0.008)	−0.004[a] (0.008)
Hlevel_dummy	−0.041[a] (0.025)	0.004[c] (0.014)	−0.012[c] (0.015)	−0.013[a] (0.015)	−0.013[a] (0.015)
Dmedical_quality	N/A[c]	0.013[a] (0.015)	0.013[a] (0.015)	0.076[a] (0.022)	0.052[a] (0.012)
Dservice_attitude	N/A	0.157[a] (0.014)	0.158[a] (0.014)	0.222[a] (0.019)	0.233[a] (0.019)
Hreputation	N/A	N/A	0.001[a] (0.001)	0.001 (0.001)	0.001 (0.001)
Severity_diseases	N/A	N/A	−0.024[a] (0.011)	−0.004[d] (0.001)	−0.004[d] (0.001)
Dmedical_quality×Hreputation	N/A	N/A	N/A	0.002[d] (0.001)	0.008[a] (0.002)
Dservice_attitude×Hreputation	N/A	N/A	N/A	−0.004[d] (0.002)	−0.007[a] (0.002)
Dmedical_quality×Severity_diseases	N/A	N/A	N/A	−0.036[a] (0.014)	−0.039[a] (0.014)
Dservice_attitude×Severity_diseases	N/A	N/A	N/A	−0.044[a] (0.012)	−0.063[a] (0.012)
Dmedical_quality×Hreputation×Severity_diseases	N/A	N/A	N/A	N/A	−0.009 (0.014)
Dservice_attitude×Hreputation×Severity_diseases	N/A	N/A	N/A	N/A	0.004 (0.012)
Log likelihood	−4790.53	−4531.46	−4233.40	−3610.23	−3577.21
Pseudo-R^2	0.014	0.015	0.018	0.020	0.021

[a] Significant at .001.

[b] Significant at .05.

[c] N/A: not applicable.

[d] Significant at .01.

Table 4. Results for the increment of outpatient care service demands. Ordinary least squares regression was used to obtain results.

Variables	Model 1
Constant	−0.188[a] (0.018)
Dtitle_dummy1	0.399[a] (0.015)
Dtitle_dummy2	0.143[a] (0.014)
Hlevel_dummy	0.039[b] (0.023)
Dmedical_quality	0.341[c] (0.110)
Dservice_attitude	1.859[a] (0.098)
Hreputation	0.018[a] (0.002)
Severity_diseases	0.029 (0.019)
Dmedical_quality×Hreputation	0.051[c] (0.020)
Dservice_attitude×Hreputation	0.050[c] (0.018)
Dmedical_quality×Severity_diseases	0.319[b] (0.148)
Dservice_attitude×Severity_diseases	−0.325[b] (0.131)
Adjusted R^2	0.785

[a] Significant at .001.

[b] Significant at .05.

[c] Significant at .01.

Robustness Check

In our study, it was found that many doctors did not receive any reviews from patients, which may have caused bias in our findings. A small increment in treatment experiences will not change the doctor’s reputation too much [17]. To check the robustness of our findings, we only included doctors whose increments of treatment experiences were equal to or greater than 1, 5, and 10. The results of the sensitivity analyses are shown in Table 5. As the results are almost identical to those shown in Table 3, our findings are deemed quite robust.

Table 5. Robustness check results. General linear model regression was used to obtain results.

Variables	The increment of treatment experiences ≥1; n=4461	The increment of treatment experiences ≥5; n=2462	The increment of treatment experiences ≥10; n=1651
Constant	0.351[a] (0.014)	−0.055 (0.085)	−0.319 (0.275)
Dtitle_dummy1	−0.029[a] (0.002)	−0.037[a] (0.003)	−0.043[a] (0.003)
Dtitle_dummy2	−0.017[a] (0.002)	−0.028[a] (0.003)	−0.033[a] (0.003)
Hlevel_dummy	−0.106[a] (0.006)	−0.178[a] (0.010)	−0.179[a] (0.012)
Dmedical_quality	0.014[b] (0.010)	0.108[a] (0.066)	0.489[a] (0.154)
Dservice_attitude	0.028[b] (0.016)	0.299[a] (0.155)	0.718[c] (0.288)
Hreputation	−0.012[a] (0.001)	−0.018 (0.008)	0.112 (0.027)
Severity_diseases	−0.121[a] (0.016)	−0.252[a] (0.061)	−0.170[c] (0.030)
Dmedical_quality×Hreputation	0.007[c] (0.002)	0.001[c] (0.000)	0.055[b] (0.023)
Dservice_attitude×Hreputation	−0.006[b] (0.002)	−0.017[c] (0.002)	−0.032[c] (0.018)
Dmedical_quality×Severity_diseases	−0.021[c] (0.014)	−0.147[c] (0.080)	−0.234[a] (0.063)
Dservice_attitude×Severity_diseases	−0.004[a] (0.001)	−0.414[a] (0.101)	−0.206[c] (0.103)
Log likelihood	−4021.50	−5112.12	−5825.11

[a] Significant at .001.

[b] Significant at .01.

[c] Significant at .05.

Discussion

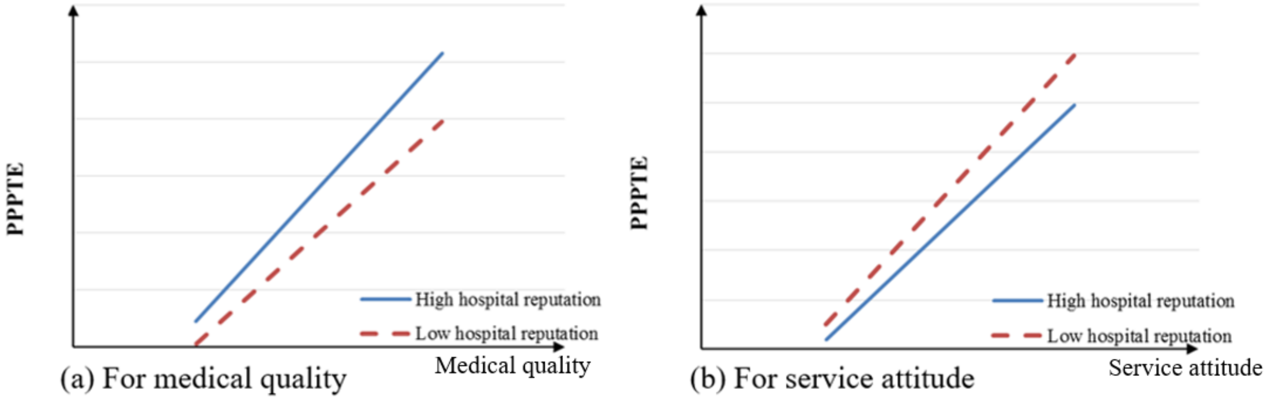
Principal Findings

This study provides valuable insights into the impact factors of sharing patient reviews in online health care communities. We study the impact of individual reputation, organizational reputation, and consumer characteristics on patient propensity to post treatment experiences and the moderating effects of organizational reputation and patient characteristics. From our results, most of the hypotheses are supported.

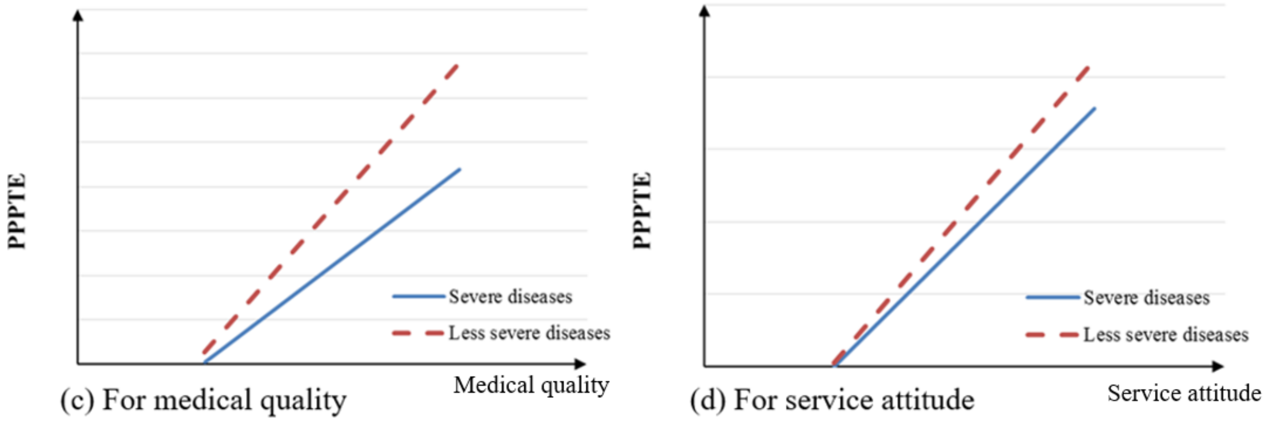
Our findings suggest that both medical quality and service attitude positively impact patient propensity to post treatment experiences, which is consistent with the expectation-confirmation theory [31] and existing literature [17,34]. When patients choose doctors with a high reputation, they hold higher expectations and are more likely to be disappointed after receiving treatment services. Similarly, when patients choose doctors who work in hospitals with high reputation, they again have high expectations, decreasing the possibility of confirmation between expectation and perceived performance.

Our results provide further evidence for the theory of psychological choice [18], as we explore the moderating effects of environmental factors in health care. The interaction effects are illustrated in Figure 4. Hospitals with high reputations can minimize patients' perceived risks and increase trust in doctors' reputations, which is confirmed in our results; a hospital's reputation positively moderates the relationship between a doctor's medical quality and patient propensity to post treatment experiences. However, a hospital's reputation negatively moderates the relationship between doctors' service attitude and patient propensity to post treatment experiences. In China, patients prefer to choose hospitals with high reputation and make appointments with doctors who also have a high reputation. As a result, hospitals with a high reputation for both their facilities and doctors are under tremendous pressure and constantly overloaded; for this reason, patients have difficulty in making appointments with these hospitals and doctors. Due to excessive number of patients, such hospitals often have low service attitudes. Therefore, patients often have low expectations about service attitude when choosing doctors who work in hospitals with high reputation.

Figure 4. The moderating effects of hospitals' reputation and disease severity. PPPTE: patient propensity to post treatment experiences.



(a) For medical quality (b) For service attitude



(c) For medical quality (d) For service attitude

Strengths and Limitations

In this study, we examine the theory of psychological choice [18] by researching the moderating effects of consumer characteristics. The results are consistent with our a priori hypothesis, and the possible explanations are as follows. First, medical services employing specialized knowledge and technology are difficult for patients to evaluate, especially for severe diseases. Consequently, following the decrease in certainty of disconfirmation between expectation and perceived quality, patient propensity to post treatment experiences decreases. Second, patients with severe diseases are more concerned about their recovery and health, instead of posting reviews. Third, for the protection of privacy and information security, patients with severe diseases may not post reviews about their treatment experiences.

Our study makes several contributions to the literature. First, this is one of the earliest in-depth studies to analyze the role of reputation in patient propensity to post treatment experiences. Prior studies have focused on the relationship between reputation and sellers' sales in both product fields [6] and health care [19]. In our study, we explore how reputation influences patient propensity to post treatment experiences, which is an important predictor of a seller's future performance [6]. Our study broadens the knowledge base on how patients think about doctors and enriches the literature on the reputation and motivation of online reviews in health care.

Second, this study contributes to the existing literature on reputation by researching the role of individual reputation, organizational reputation, and interaction effects. Prior studies have only considered reputation at one level, either individual [54] or organizational [43]. In health care, Wu and Lu [17] researched the impact of individual reputation on patients' propensity to post reviews, but they failed to consider organizational reputation. Both individual and organizational reputations work effectively to determine consumer behavior, especially for medical services with high information asymmetry [55]. In the health care field, doctors are affiliated with hospitals. Patients often place great importance on a hospital's reputation (environment of signal delivery), which must be considered in health care. Our study helps understand the role of doctors' (individual) reputation, hospital (organizational) reputation, and their interaction effects on patient propensity to post treatment experiences.

Third, we enrich the existing literature on the impacts of consumer characteristics on consumer behavior. Consumer characteristics have been recognized by researchers as one of the most influential factors for different consumer behaviors [19,27]. Among these patient characteristics, disease severity is extremely important. However, there is not much current investigation into the impact of patient characteristics on their behavior, which is measured by their propensity to post reviews about treatment experiences. Our research provides an empirical analysis of the theory of psychological choice by examining the moderating effects of disease severity on whether patients post their experiences.

This study also has significant practical implications. First, our results show that when patients decide whether to post treatment experience reviews, service attitude works more effectively than medical quality. Our findings also suggest that doctors need to pay more attention to their service attitude than ever before. When people have diseases, they become vulnerable and seek emotional support from doctors. Moreover, contradictions and disputes between doctors and patients have intensified, which has reached an unprecedented level in recent years, requiring doctors to improve their service attitudes. Second, not only do we find that doctors' reputation has a positive impact on the number of reviews posted but also the hospital's reputation; thus, to encourage more patients to post reviews online, doctors must take the impact of the hospital's reputation into consideration. For example, doctors can move to other hospitals with higher reputation. Third, disease severity mitigates the relationship between doctor reputation and patient propensity to post treatment experiences. Compared with doctors who treat severe diseases, doctors who treat less severe diseases should pay closer attention to their online reputation. As their online reputation increases, doctors who treat less severe diseases receive a greater number of patient reviews than those who treat severe diseases.

Our study has several limitations. First, we include one online service, the WeDoctor website. Although improving the internal validity, this choice may reduce the generalizability of our findings. Other contexts should be examined in future studies. Second, we did not collect patient-level data; because of this, we could not measure demographic characteristics and specific disease severity for each patient. Future research can improve our findings by collecting data at the patient level. Third, we did not analyze the content of treatment experience reviews. These new treatment experiences may reflect different feelings and play different roles and should be investigated in future studies. Last but not least, future studies should adopt a longitudinal approach to improve our findings by addressing potential endogeneity issues and dynamic effects.

Acknowledgments

This work was supported by the National Natural Science Foundation of China grant number 72001087.

Conflicts of Interest

None declared.

References

1. Xu Y, Armony M. The effect of online reviews on physician demand: a structural model of patient choice. SSRN J 2016:-epub ahead of print. [doi: 10.2139/ssrn.2778664]
2. More Than 80% of Patients Search for Healthcare Information Before Travelling to Hospitals. Ifeng. 2011. URL: http://news.ifeng.com/gundong/detail_2011_06/24/7232672_0.shtml [accessed 2020-12-01] [WebCite Cache ID http://news.ifeng.com/gundong/detail_2011_06/24/7232672_0.shtml]
3. Dellarocas C. The digitization of word of mouth: promise and challenges of online feedback mechanisms. Manag Sci 2003 Oct;49(10):1407-1424. [doi: 10.1287/mnsc.49.10.1407.17308]
4. Bloom PN, Reve T. Transmitting signals to consumers for competitive advantage. Bus Horiz 1990 Jul;33(4):58-66. [doi: 10.1016/0007-6813(90)90059-k]
5. Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. Int J Med Inform 2018 Feb;110:77-89. [doi: 10.1016/j.ijmedinf.2017.11.009] [Medline: 29331257]
6. Ye Q, Xu M, Kiang M, Wi W, Sun F. In-depth analysis of the seller reputation and price premium relationship: a comparison between eBay us and Taobao China. J Electro Comm Res 2013;14:1 [FREE Full text]
7. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. Int J Med Inform 2017 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
8. Chintagunta PK, Gopinath S, Venkataraman S. The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets. Mark Sci 2010 Sep;29(5):944-957. [doi: 10.1287/mksc.1100.0572]
9. Yu H, Wang Y, Wang JN, Chiu YL, Gao M. Causal effect of honorary titles on physicians' service volumes in online health communities: retrospective study. J Med Internet Res 2020;22:8527 preprint. [doi: 10.2196/preprints.18527]
10. Forman C, Ghose A, Wiesenfeld B. Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. Inf Syst Res 2008 Sep;19(3):291-313. [doi: 10.1287/isre.1080.0193]
11. Chen Y, Xie J. Online consumer review: word-of-mouth as a new element of marketing communication mix. Manage Sci 2008 Mar;54(3):477-491. [doi: 10.1287/mnsc.1070.0810]
12. Yang J, Mai E. Experiential goods with network externalities effects: an empirical study of online rating system. J Bus Res 2010 Sep;63(9-10):1050-1057. [doi: 10.1016/j.jbusres.2009.04.029]
13. Whyte Jr WH. The web of word of mouth. Fortune 1954;50:140.
14. Zhu P, Shen J, Xu M. Patients' willingness to share information in online patient communities: questionnaire study. J Med Internet Res 2020;22:e16546. [doi: 10.2196/preprints.16546]
15. Dichter E. How word-of-mouth advertising works. Harvard Bus Rev 1966;44:147. [doi: 10.4135/9781452229669.n3968]

16. de Matos CA, Rossi CA. Word-of-mouth communications in marketing: a meta-analytic review of the antecedents and moderators. J Acad Mark Sci 2008 Sep 23;36(4):578-596. [doi: 10.1007/s11747-008-0121-1]
17. Wu H, Lu N. How your colleagues' reputation impact your patients' odds of posting experiences: evidence from an online health community. Electronic Comm Res Appl 2016 Mar;16:7-17. [doi: 10.1016/j.elerap.2016.01.002]
18. Hansen F. Psychological theories of consumer choice. J Consum Res 1976 Dec;3(3):117. [doi: 10.1086/208660]
19. Lu N, Wu H. Exploring the impact of word-of-mouth about Physicians' service quality on patient choice based on online health communities. BMC Med Inform Decis Mak 2016 Nov 26;16(1):151 [FREE Full text] [doi: 10.1186/s12911-016-0386-0] [Medline: 27888834]
20. Robin DiMatteo M, Prince LM, Taranta A. Patients' percentions of physicians' behavior. J Commun Health 1979 Jun;4(4):280-290. [doi: 10.1007/bf01319022]
21. Ba S, Wang L. Digital health communities: the effect of their motivation mechanisms. Decis Supp Sys 2013 Nov;55(4):941-947. [doi: 10.1016/j.dss.2013.01.003]
22. Si Y, Wu H, Liu Q. Factors influencing doctors' participation in the provision of medical services through crowdsourced health care information websites: elaboration-likelihood perspective study. JMIR Med Inform 2020 Jun 29;8(6):e16704 [FREE Full text] [doi: 10.2196/16704] [Medline: 32597787]
23. Yan L, Tan Y. Feeling blue? Go online: an empirical study of social support among patients. Inf Syst Res 2014 Dec;25(4):690-709. [doi: 10.1287/isre.2014.0538]
24. Lu W, Wu H. How online reviews and services affect physician outpatient visits: content analysis of evidence from two online health care communities. JMIR Med Inform 2019 Dec 2;7(4):e16185 [FREE Full text] [doi: 10.2196/16185] [Medline: 31789597]
25. Han X, Li B, Zhang T, Qu J. Factors associated with the actual behavior and intention of rating physicians on physician rating websites: cross-sectional study. J Med Internet Res 2020;22:4417. [doi: 10.2196/preprints.14417]
26. Ni J, Sun B. A Dynamic Game of Doctors' Participation in Online Health Information Platform. In: Workshop on Health IT and Economics. 2010 Presented at: HealthIT'10; May 16-19, 2010; Maryland, USA.
27. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: an empirical analysis of a national cancer-related survey. Decis Support Sys 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
28. Bansal G, Zahedi F, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decis Supp Syst 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
29. Oliver RL. Cognitive, affective, and attribute bases of the satisfaction response. J Consum Res 1993 Dec;20(3):418. [doi: 10.1086/209358]
30. Swan JE, Trawick IF. Disconfirmation of expectations and satisfaction with a retail service. J Retailing 1981:49-67.
31. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. J Mark Res 2018 Nov 28;17(4):460-469. [doi: 10.1177/002224378001700405]
32. Rodhom C, Tolman EC. Purposive behavior in animals and men. Am J Psychol 1950 Apr;63(2):305. [doi: 10.2307/1418946]
33. Schindler RM, Bickart B. Published word of mouth: Referable, consumer-generated information on the Internet. Lawrence Erlbaum Associates Inc 2005:35-61.
34. Liu Y. Word of mouth for movies: its dynamics and impact on box office revenue. J Mark 2018 Oct 2;70(3):74-89. [doi: 10.1509/jmkg.70.3.074]
35. Khalifa M, Liu V. Satisfaction with internet-based services: the role of expectations and desires. Int J Electronic Comm 2014 Dec 23;7(2):31-49. [doi: 10.1080/10864415.2002.11044267]
36. Jung CG. A visionary rumour. J Analytical Psychol 1959 Jan;4(1):5-19. [doi: 10.1111/j.1465-5922.1959.00005.x]
37. Anderson EW. Customer satisfaction and word of mouth. J Serv Res 2016 Jun 29;1(1):5-17. [doi: 10.1177/109467059800100102]
38. Dellarocas C, Gao G, Narayan R. Are consumers more likely to contribute online reviews for hit or niche products? J Manag Inform Syst 2014 Dec 8;27(2):127-158. [doi: 10.2753/mis0742-1222270204]
39. Amblee N, Bui T. Can brand reputation improve the odds of being reviewed on-line? Int J Electron Comm 2014 Dec 8;12(3):11-28. [doi: 10.2753/jec1086-4415120302]
40. Wu H, Deng Z, Wang B, Wu T. Online service qualities in the multistage process and patients' compliments: a transaction cycle perspective. Inf Manag 2020 Jul;57(5):103230. [doi: 10.1016/j.im.2019.103230]
41. Wu H, Deng Z. Knowledge collaboration among physicians in online health communities: a transactive memory perspective. Int J Inf Manag 2019 Dec;49:13-33. [doi: 10.1016/j.ijinfomgt.2019.01.003]
42. Bailey AA. Non-fulfillment of promotional deals: the impact of gender and company reputation on consumers' perceptions and attitudes. J Retailing Consum Serv 2005 Jul;12(4):285-295. [doi: 10.1016/j.jretconser.2004.10.003]
43. Yoon E, Guffey HJ, Kijewski V. The effects of information and company reputation on intentions to buy a business service. J Bus Res 1993 Jul;27(3):215-228. [doi: 10.1016/0148-2963(93)90027-m]
44. Connelly BL, Certo ST, Ireland RD, Reutzel CR. Signaling theory: a review and assessment. J Manag 2010 Dec 20;37(1):39-67. [doi: 10.1177/0149206310388419]
45. Mavlanova T, Benbunan-Fich R. Counterfeit products on the internet: the role of seller-level and product-level information. Int J Electron Comme 2014 Dec 10;15(2):79-104. [doi: 10.2753/jec1086-4415150203]

46. Ulrich I. The effect of consumer multifactorial gender and biological sex on the evaluation of cross-gender brand extensions. Psychol Mark 2013 Jul 23;30(9):794-810. [doi: 10.1002/mar.20646]
47. Zhu F, Zhang X. Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. J Mark 2010 Mar;74(2):133-148. [doi: 10.1509/jm.74.2.133]
48. Cooil B, Keiningham TL, Aksoy L, Hsu M. A longitudinal analysis of customer satisfaction and share of wallet: investigating the moderating effect of customer characteristics. J Mark 2007 Jan;71(1):67-83. [doi: 10.1509/jmkg.71.1.67]
49. Anderson S, Pearo LK, Widener SK. Drivers of service satisfaction. J Serv Res 2008 Mar 13;10(4):365-381. [doi: 10.1177/1094670508314575]
50. Ruo B, Rumsfeld JS, Hlatky MA, Liu H, Browner WS, Whooley MA. Depressive symptoms and health-related quality of life. J Am Med Assoc 2003 Jul 9;290(2):215. [doi: 10.1001/jama.290.2.215]
51. Buller MK, Buller DB. Physicians' communication style and patient satisfaction. J Health Soc Behav 1987 Dec;28(4):375. [doi: 10.2307/2136791]
52. Berry S, Levinsohn J, Pakes A. Automobile prices in market equilibrium. Econometrica 1995 Jul;63(4):841. [doi: 10.2307/2171802]
53. The Causes of Residents' Death. The State Statistics Bureau, the Chinese Health Statistics Yearbook. 2014. URL: http://www.stats.gov.cn/tjsj/ndsj/2014/indexch.htm [accessed 2020-12-01]
54. Ba S, Pavlou PA. Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. MIS Q 2002 Sep;26(3):243. [doi: 10.2307/4132332]
55. Liu X, Guo X, Wu H, Wu T. The impact of individual and organizational reputation on physicians' appointments online. Int J Electron Comme 2016 Jun 20;20(4):551-577. [doi: 10.1080/10864415.2016.1171977]

Edited by G Eysenbach; submitted 14.10.19; peer-reviewed by Y Song, C Xu, JN Wang, Y Wang; comments to author 24.08.20; revised version received 03.10.20; accepted 11.11.20; published 11.12.20.

Please cite as:
Wang Y, Wu H, Lei X, Shen J, Feng Z
The Influence of Doctors' Online Reputation on the Sharing of Outpatient Experiences: Empirical Study
J Med Internet Res 2020;22(12):e16691
URL: http://www.jmir.org/2020/12/e16691/
doi:10.2196/16691
PMID:33306028

Original Paper

COVID-19 as 'Game Changer' for the Physical Activity and Mental Well-Being of Augmented Reality Game Players During the Pandemic: Mixed Methods Survey Study

Louise A Ellis[1,2*], PhD; Matthew D Lee[3*], MS, RN; Kiran Ijaz[4], PhD; James Smith[1], PhD; Jeffrey Braithwaite[1,2], PhD; Kathleen Yin[4], PhD

[1]Centre for Healthcare Resilience and Implementation Science, Australian Institute of Health Innovation, Macquarie University, North Ryde, Australia
[2]NHMRC Partnership Centre in Health System Sustainability, Australian Institute of Health Innovation, Macquarie University, North Ryde, Australia
[3]School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
[4]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, North Ryde, Australia
[*]these authors contributed equally

Corresponding Author:
Louise A Ellis, PhD
Centre for Healthcare Resilience and Implementation Science
Australian Institute of Health Innovation
Macquarie University
Level 6, Talavera Road
North Ryde, 2109
Australia
Phone: 61 298502484
Email: louise.ellis@mq.edu.au

Abstract

Background: Location-based augmented reality (AR) games, such as Pokémon GO and Harry Potter: Wizards Unite, have been shown to have a beneficial impact on the physical activity, social connectedness, and mental health of their players. In March 2020, global social distancing measures related to the COVID-19 pandemic prompted the AR games developer Niantic Inc to implement several changes to ensure continued player engagement with Pokémon GO and Harry Potter: Wizards Unite. We sought to examine how the physical and mental well-being of players of these games were affected during the unprecedented COVID-19 restriction period as well as how their video game engagement was affected.

Objective: The aims of this study were to examine the impact of COVID-19–related social restrictions on the physical and mental well-being of AR game players; to examine the impact of COVID-19–related social restrictions on the use of video games and motivations for their use; and to explore the potential role of AR games (and video games in general) in supporting well-being during COVID-19–related social restrictions.

Methods: A mixed methods web-based self-reported survey was conducted in May 2020, during which COVID-19–related social restrictions were enforced in many countries. Participants were recruited on the web via four subreddits dedicated to Pokémon GO or Harry Potter: Wizards Unite. Data collected included quantitative data on demographics, time spent playing video games, physical activity, and mental health; qualitative data included motivations to play and the impact of video games on mental health during COVID-19 lockdown.

Results: We report results for 2004 participants (1153/1960 male, 58.8%, average age 30.5 years). Self-reported physical activity during COVID-19–related social restrictions significantly decreased from 7.50 hours per week on average (SD 11.12) to 6.50 hours (SD 7.81) (P<.001). More than half of the participants reported poor mental health (925/1766, 52.4%; raw World Health Organization–5 Well-Being Index score <13). Female gender, younger age, and reduced exercise were significant predictors of poor mental health. Participants reported a significant increase in video game play time from 16.38 hours per week on average (SD 19.12) to 20.82 hours (SD 17.49) (P<.001). Approximately three quarters of the participants (n=1102/1427, 77.2%) reported that playing video games had been beneficial to their mental health. The changes made to Pokémon GO and Harry Potter: Wizards Unite were very well received by players, and the players continued to use these games while exercising and to maintain social connection. In addition to seeking an escape during the pandemic and as a form of entertainment, participants reported that they used video games for emotional coping and to lower stress, relax, and alleviate mental health conditions.

Conclusions: AR games have the potential to promote physical and mental health during the COVID-19 pandemic. Used by populations under isolation and distress, these games can improve physical and mental health by providing virtual socialization, sustained exercise, temporal routine, and mental structure. Further research is needed to explore the potential of AR games as digital behavioral interventions to maintain human well-being in the wider population.

(J Med Internet Res 2020;22(12):e25117) doi:10.2196/25117

KEYWORDS

COVID-19; Pokémon GO; Harry Potter: Wizards Unite; augmented reality games; physical activity; mental health; well-being

Introduction

COVID-19 was first reported in Wuhan, China, on December 31, 2019 [1,2], and it has since escalated to a global pandemic. COVID-19 is transmitted between humans in close proximity; therefore, physical (or social) distancing is a key measure for reducing its spread [3]. By April 2020, most countries worldwide had introduced quarantine measures and travel bans, cancelled social events, and closed public services to contain COVID-19 [4]. With the introduction of stay-at-home lockdown and quarantine measures, electronic video game playing reached an all-time high [5]. Reputedly, 82% of global consumers played video games and watched video game content during the height of the lockdown period during the COVID-19 pandemic [5]. Increased web-based gaming was viewed as complementary to public efforts to promote physical distancing [6]. Most notably, the World Health Organization (WHO) partnered with the gaming industry in March 2020 to launch the campaign #PlayApartTogether to encourage people to stay at home, play video games, and practice physical distancing [7].

Although web-based gaming has played a supporting role during the COVID-19 pandemic in maintaining physical distancing, concerns have been raised about encouraging video gaming [6]; previous research has linked excessive gaming with poor mental health, sleep problems, and physical inactivity [8]. Protracted periods of isolation, technology-based activity, and limited social interaction can also solidify unhealthy lifestyle patterns, intensify technology-related disorders, and potentially lead to difficulties of readaptation after the COVID-19 pandemic has ended [6].

For this reason, “healthier” gaming options that promote increased physical activity and social connection [6,9,10] are being encouraged. Location-based augmented reality (AR) games developed by Niantic Inc, such as Pokémon GO and Harry Potter: Wizards Unite [11], are of particular interest because they are designed to increase physical activity [12,13] and have also been found to increase players’ social connectedness [14] and mental well-being [12,15]. These games require players to explore outdoor public spaces and engage in social events as part of the normal play experience, using cellular, Wi-Fi, and GPS networks to determine the player’s approximate location. This in turn affects the player’s ability to interact with game features and other players.

With government-issued restrictions during the COVID-19 pandemic encouraging—or requiring—individuals to stay at home, playing AR games in the manner in which they were originally designed has become challenging [10,16]. Niantic Inc implemented several in-game changes to Pokémon GO and Harry Potter: Wizards Unite to prevent these restrictions from having deleterious effects on the player’s experience; therefore, these games are easier to play at home and in social isolation [17]. These changes ensured that Pokémon GO and Harry Potter: Wizards Unite continued to be popular [16]. However, these changes did not necessarily translate to improved physical and mental well-being for players, as many gameplay features designed to promote physical activity and social activity were removed or altered [16]. This study sought to examine the physical and mental well-being of players of Pokémon GO and Harry Potter: Wizards Unite during the COVID-19 pandemic as well as how players viewed the roles played by AR games and video games in general in their mental health during the pandemic.

As such, our aims were (1) to examine the impact of COVID-19–related social restrictions on the physical and mental well-being of AR game players; (2) to examine the impact of COVID-19–related social restrictions on use of video games and motivations for their use; and (3) to explore the potential role of AR games (and video games in general) in supporting players’ well-being during COVID-19–related social restrictions.

Methods

Overview

We conducted a mixed methods web-based survey during a period in which many countries were under social restrictions due to COVID-19 (May 15-29, 2020, inclusive) [4]. This survey is part of a larger study examining Pokémon GO and Harry Potter: Wizards Unite use and player experience during the COVID-19 pandemic. Ethical approval was obtained from the Human Research Ethics Committee for Medical Sciences of Macquarie University (Reference No: 52019601512435).

Sample and Recruitment

Participant recruitment was conducted on the web via four subreddits dedicated to Pokémon GO or Harry Potter: Wizards Unite (r/WizardsUnite, r/PokemonGO, r/harrypotterwu, and r/TheSilphRoad). A recruitment post was pinned to the top of each of subreddit from May 15-29, 2020, directing individuals to the participant information page via Qualtrics [18]. Participants were informed that participation was voluntary, that they could withdraw at any time prior to submitting their final responses, and that to be eligible, they needed to be ≥18 years of age and have played the English versions of Pokémon GO or Harry Potter: Wizards Unite for at least a week. Those who consented to participate were directed to the web-based survey, which took 15 to 20 minutes to complete. No incentives were offered for taking part in the survey. Participants were informed not to provide any identifiable personal information. The IP addresses of the participants were recorded by Qualtrics to identify repeat visits. Duplicate entries were not permitted after a participant had submitted the survey.

The Survey

The survey included a total of 40 questions, both quantitative and qualitative. Demographic information was collected from the participants, including age, gender, and country of residence. Responses from quantitative questions relating to video game use, exercise, and mental well-being are reported here, as well as responses from two qualitative questions regarding motivation to play and the impact of games on mental health (Multimedia Appendix 1).

Video Game Use

Participants were asked whether they identified as a “hardcore gamer,” a “casual gamer,” or a “midcore gamer” to identify their usual play patterns and levels of engagement [19]. Hardcore gamers were defined as those who invest a significant amount of their recreational time and resources into gaming and who also have more knowledge and skills related to games [19-21]. On the other hand, casual gamers were defined as those who play games casually, in short sessions, or infrequently [22,23]. Midcore gamers fall in between these categories and include people who regularly play video games but are not “super serious” gamers [23]. Participants were also asked for their typical video game playing frequency prior to the COVID-19 pandemic (number of hours per day and days per week) and during COVID-19 lockdown (number of hours per day and days per week). In addition, participants were asked whether they continued to play either Pokémon GO or Harry Potter: Wizards Unite during the lockdown (Yes/No).

Physical Activity

Respondents were asked to report their typical exercise frequency prior to the COVID-19 pandemic (number of hours per day and days per week) and during COVID-19 lockdown (number of hours per day and days per week).

Mental Well-being

The World Health Organization–5 Well-Being Index (WHO-5) [24] is a 5-item measure of current mental well-being. The WHO-5 consists of five items assessing positive mood, vitality, and general interest over the past two weeks, scored on a 6-point Likert scale from 0 (at no time) to 5 (all of the time). Scores are summed to create a total “raw score” (range 0-25), with lower scores indicating impaired emotional well-being [24]. A total “percentage score” is obtained by multiplying the total raw score by four. A raw score <13 (or <50%, or a score of 0 or 1 on any of the items) is considered to be an indicator of impaired mental well-being and of likely depression [25,26].

Qualitative Questions

This paper focuses on responses to two qualitative questions: (1) “What motivates you to play Pokémon GO or Harry Potter: Wizards Unite during the COVID-19 shutdown?” and (2) “How has playing video games affected your mental health during the shutdown?”

Data Analysis

Quantitative survey data were analyzed using SPSS 25.0 (IBM Corporation). Reported total hours of participation in gaming and exercise per week were calculated by multiplying the number of hours by the number of days they were performed. Differences between male and female participants and between self-identified player types were examined using chi-square analysis or multinomial logistic regression for categorical dependent variables or analysis of variance (ANOVA) for continuous dependent variables. Repeated-measures ANOVA was also used to assess mean differences in hours of participation in gaming and exercise per week by gender and player type over time (ie, pre–COVID-19 and during COVID-19 lockdown). Finally, binary logistic regression was used to determine factors associated with impaired mental well-being. Here, game play and exercise change scores (pre–COVID-19 minus during COVID-19 lockdown) were incorporated in the analysis. Due to the large sample size, the significance level was set at $P<.01$ for all analyses.

Qualitative responses were analyzed via thematic analysis using NVivo (QSR International) [27]. Thematic analysis using a 6-step process [28] was undertaken independently by two researchers (LAE and MDL). Codes were developed using an iterative process to finalize and map important themes and ensure consistency with the data [29]. The broader research team were included throughout each stage of the analysis process to help resolve any differences, with frequent discussions concerning themes. The qualitative findings were used to triangulate and build upon the quantitative data.

Results

Demographic Results

In total, 2165 participants responded to the survey. The sample was reduced to 2004 after excluding participants with more than 30% missing survey data. In the sample, most participants were male (1153/1960, 58.8%) and aged between 25 and 34 years (n=985/1998, 49.3%, M=30.5, range:18-99). Participants were drawn from 66 different countries, although almost half were from the United States (987/1987, 49.7%). The characteristics of the survey respondents are presented in Table 1. Only countries with more than 50 respondents are shown.

Table 1. Demographic characteristics of survey respondents (N=2004), n (%). Values may not equal total N due to missing demographic responses.

Characteristic	Value
Gender	
Female	807 (41.2)
Male	1153 (58.8)
Age (years)	
18-24	523 (26.2)
25-34	985 (49.3)
35-44	329 (16.5)
≥45	161 (8.1)
Country	
Australia	77 (3.9)
Canada	146 (7.5)
Germany	90 (4.5)
The Netherlands	52 (2.6)
United Kingdom	177 (8.9)
United States	987 (49.7)
Other	458 (23.1)
Self-identified player type	
Casual gamer	905 (46.6)
Hardcore gamer	392 (20.2)
Midcore gamer	646 (33.2)

Self-Identified Player Type

When asked to self-identify as a hardcore, casual, or midcore gamer, most respondents considered themselves casual gamers (905/1943, 46.6%), followed by midcore gamers (646/1943, 33.2%), and hardcore gamers (392/1943, 20.2%). Hardcore gamers were significantly more likely to be male (324/385, 84.2%) than female (61/385, 16.0%) (χ^2_{1901}=160.23, P<.001), and significantly younger than casual gamers (χ^2_{1293}=53.97, P<.001).

Quantitative Results

Impact of the COVID-19 Pandemic on Physical Activity

Prior to the COVID-19 pandemic, participants reported exercising for an average of 7.50 hours per week (SD 11.12). At this time, hardcore players (mean 9.07 hours, SD 12.92) exercised significantly more than casual gamers (mean 7.19 hours, SD 11.10; $F_{1,1129}$=6.14, P=.01), and men (mean 8.31 hours, SD 13.00) exercised significantly more than women (mean 6.35 hours, SD 7.88; $F_{1,1708}$=12.88, P<.001). During the COVID-19 lockdown, participants decreased their exercise to an average of 6.50 hours per week (SD 7.81). Repeated measures ANOVA showed that this reduction in exercise from before to during the lockdown was significant ($F_{1,1058}$=15.79, P<.001). A significant player type–by-time interaction was also identified, with hardcore gamers exhibiting a greater reduction in exercise during the COVID-19 lockdown than casual gamers ($F_{1,1058}$=6.58, P=.01). There was no significant gender-by-time interaction. We also identified a significant positive correlation between total hours of gaming and exercise per week both during COVID-19 lockdown (r=0.198, P<.001) and prior to lockdown (r=0.257, P<.001).

Mental Well-being During COVID-19 Lockdown

Over half of the participants reported poor mental health (925/1766, 52.4%; raw WHO-5 score <13) at the time the survey was conducted. A multivariate binary logistic regression analysis was conducted with age, gender, self-identified gaming type, difference in video game play, and difference in exercise entered as predictors. The model identified being female (P<.001), being younger (P<.001), and reduced exercise (P=.008) as significant predictors of poor mental health during the COVID-19 lockdown. At a significance level of P<.01, self-identified player type (P=.22) and difference in video game play (P=.04) were not significant predictors of poor mental health (Multimedia Appendix 2).

Impact of COVID-19 Lockdown on Video Game Use

Three in four participants (1261/1675, 75.3%) reported that during the COVID-19 lockdown, they played video games "a little more" to "a lot more." Of these 1675 participants, 20% (n=335) reported "no change," and only 4.7% (n=79) reported playing "a little less" to "a lot less." Prior to the COVID-19 pandemic, participants played video games for an average of 16.38 hours per week (SD 19.12). At this time, hardcore gamers (mean 20.24 hours, SD 22.47) played significantly more than casual gamers (mean 13.97, SD 17.32; $F_{1,1276}$=29.27, P<.001), and men (mean 17.90, SD 20.77) played significantly more than women (mean 14.26, SD 16.34; $F_{1,1929}$=17.03, P<.001). During the COVID-19 lockdown, participants increased their video game playing time to an average of 20.82 hours per week (SD 17.49). Repeated measures ANOVA indicated that the difference in video game playing time between the pre-lockdown and lockdown periods was significant ($F_{1,1260}$=55.12, P<.001). Notably, a significant player type–by-time interaction was found; casual gamers increased their video game playing time during COVID-19 lockdown (mean 20.00 hours, SD 15.35) to almost match that of hardcore gamers (mean 21.84 hours, SD 19.91; $F_{1,1260}$=16.83, P<.001). Similarly, a significant gender-by-time interaction was identified; women increased their video game playing time during COVID-19 lockdown (mean 20.91 hours, SD 17.43) to equal that of men (mean 20.82 hours, SD 17.71; $F_{1,1897}$=17.97, P<.001). Further, virtually all participants (1962/1975, 99.3%) reported that they were continuing to play either Pokémon GO or Harry Potter: Wizards Unite during lockdown.

Qualitative Results

Motivation to Play AR Games During COVID-19 Lockdown

For the first qualitative question, "What motivates you to play Pokémon GO or Harry Potter: Wizards Unite during the COVID-19 shutdown?" several key themes emerged from the participants' responses (Multimedia Appendix 3). The most prominent themes were entertainment, achievements and challenges, in-game modifications, and exercise.

Entertainment

Many participants (506/1527, 33.1%) indicated that they played Pokémon GO or Harry Potter: Wizards Unite during the pandemic because these games were "fun" and "entertaining." Some participants noted that these games provided them with "something to do" during the lockdown, as there were few alternatives they could enjoy while "stuck at home" and, in some cases, under financial constraints due to being unemployed or underemployed. Other participants mentioned that COVID-19–related restrictions gave them more time to play, with one 25-year-old male participant stating that it gave him "an opportunity to come back to the game after a few years," that he "missed playing it," and that it gave him "something to do in [his] downtime."

Achievements and Challenges

Almost one-third of the participants (500/1527, 32.7%) stated that achievements and challenges were key motivating factors for playing Pokémon GO or Harry Potter: Wizards Unite. Participants stated that the "daily challenges" and "daily tasks" gave them a "sense of accomplishment." Many Pokémon GO players said that they were driven to accomplish the primary objective of the game, "catch Pokémon," as well as other objectives such as hunting for "shinies" (color-variant Pokémon) and receiving in-game rewards. Several participants specifically identified having tasks and challenges to complete in Pokémon GO and Harry Potter: Wizards Unite as providing them with beneficial temporal structure and achievable goals, given the disruptions to work or school posed by COVID-19–related restrictions. For example, one female participant aged 35 years noted that it provided "a challenge/task to complete when most of the deadlines have been removed from my life." Another female participant aged 24 years said, "I've been playing Wizards Unite during the lockdown as a way to fill time and have an activity that provides a sense of progress and achievement."

In-Game Modifications

Many respondents (343/1527, 22.5%) highlighted that they liked or even "loved" the new changes made to Pokémon GO and Harry Potter: Wizards Unite. The games became "more accessible to play from home," and the "increased rewards and events have kept [people] engaged." One 30-year-old male participant praised Niantic for their efforts: "Niantic has done well to implement features that make playing from home easier." Notably, changes such as the "Knight Bus" in Harry Potter: Wizards Unite (a function that enabled players to cooperatively face in-game challenges from home) and the opening of the "GO Battle League" in Pokémon GO (which allowed players to challenge others and compete for rankings) reduced or eliminated the need to walk outdoors. One female Harry Potter: Wizards Unite player (age 53 years) elaborated:

> *The Knight Bus completely changed game play for me…Online instructions and the ability to 'travel' to a Fortress has made the game 100% more interesting.*

Pokémon GO players also appreciated the addition of "spotlight hours, improvements to the incense and such" (male participant, age 20 years), but it was "GO Battle League" that was cited as "game-changing," with one male player (age 30 years) mentioning that the removal of walking requirements to participate had "renewed [his] interest in the game" by providing "a measurable way to make [and measure] progress."

Exercise

One-fifth of respondents (316/1527, 20.7%) noted that they used Pokémon GO or Harry Potter: Wizards Unite while exercising. Many stated that these games motivated them to exercise more, with one 24-year-old male participant mentioning that "it is the only thing that motivates me to go on walks outside" and another 33-year-old female participant saying that these games were "the only reason to step outside" and she was "actually exercising more during lockdown." For others, it gave them "something to do during exercise" (male participant, age 29 years), or during other activities such as "walking the dog" or spending time with family members. Some elaborated on their exercise routines during the COVID-19 restrictions, with a typical example below:

> *Pokémon GO is the one thing that keeps me going outdoors and moving each day. It gives me a purpose to keep walking to different parts of the city but also lets me do so while staying away from clusters of other people.* [Female participant, age 28 years]

Impact of Video Games on Mental Well-being During COVID-19

Analysis of responses to the second qualitative question indicated that more than three in four participants (1102/1427, 77.2%) believed that playing video games during COVID-19 lockdown had been beneficial to their mental health; 20.7% of the participants (295/1427) gave a neutral response, and only 2.1% (30/1427) reported a negative impact. For participants who reported a positive mental health impact, several key themes described the role of video during the COVID-19 lockdown (Multimedia Appendix 4).

Escape/Distraction

Approximately half of the participants who reported a positive mental health impact (541/1102, 49.1%) identified that video games had been helpful in "providing a much-needed escape" and "a great distraction" from the current situation. A number of participants emphasized that games were particularly helpful in how they took their "mind off the constant depressing news coverage of COVID-19." One 34-year-old female participant elaborated: "I can focus on the game instead of obsessively checking the news and stressing out." Another male participant (age 27 years) said: "[They] kept me away from watching the news all day. It was quite nice to not be surrounded by COVID news all the time."

Activity/Entertainment

Almost half of the participants (535/1102, 48.5%) stated that video games gave them "something to do," helped them to "stay busy," and kept them "entertained" during COVID-19 lockdown. Games gave them "something fun to focus on" and "helped to relieve boredom" while "being at home 24/7." Many described how games gave them "something to look forward to," with one 30-year-old male participant even stating that beyond keeping him "motivated and busy," games helped him to "get out of bed and do something."

Emotional Coping

In addition to providing an escape and serving as a way to pass the time or be entertained, many participants (469/1102, 42.6%) identified that video games had helped them "cope" and maintain "a calm and positive outlook." Playing video games allowed them to "relieve some stress" and "relax," and it "lifted [their] mood." One 30-year-old female participant wrote that video games were "possibly the only thing keeping me sane right now." Some respondents also wrote about using video games to alleviate specific mental health conditions:

> *It's kept me in a safe place mentally. While college was in session I had depression and anxiety and it's still lingering as of now. It takes my mind off of the outside world.* [Male participant, age 19 years]
>
> *I suffered from anxiety and depression prior to the shutdown, too, and used games for the same escape then.* [Female participant, age 42 years]

Social Connection

One in five respondents (219/1102, 19.9%) emphasized the importance of video games in strengthening social connections. Many indicated that they played video games with others in their household (eg, children and partner), thus "providing bonding opportunities." Multiplayer games provided an opportunity to stay in touch with friends and "still share in something together." One 30-year-old participant stated: "It has helped because it's the only social interaction I get." Some participants highlighted the importance of the local "community" these games create, noting that playing let them feel "less isolated":

> *In Pokémon GO, the local app group makes me feel part of a community, even if we cannot all meet face to face right now."* [Female participant, age 31 years]
>
> *My HWPU [Harry Potter: Wizards Unite]/ingress community connections have been vital in having someone to talk to, and learning about other people's experiences.* [Female participant, aged 54 years]

Discussion

Principal Results

The COVID-19 pandemic has significantly disrupted normal activities worldwide. In this study, we sought to examine the impact of COVID-19 on people's physical and mental well-being and on video game use. We also sought to explore the potential role of video games (and AR games in particular) in supporting the mental well-being of players during COVID-19 lockdown. We summarize our key findings in accordance with our three objectives.

1. Impact of COVID-19–Related Social Restrictions on the Physical and Mental Well-being of AR Game Players

Due to the COVID-19–related stay-at-home restrictions, exercise levels dropped significantly from an average of 7.5 hours to 6.5 hours per week during the pandemic. Despite this drop, on average, players continued to exercise well above the level of 150 minutes (2.5 hours) of moderately intense activity recommended per week by the WHO to achieve health benefits [30]. Previous research has identified that performing at least 150 minutes of exercise a week improves cardiorespiratory and muscular fitness, improves bone health, and reduces the risk of chronic disease and depression [30]. From the quantitative survey data, we also identified a significant positive relationship between total hours of participation in gaming and total hours of exercise per week, indicating that rather than hardening unhealthy lifestyle patterns [6], AR games may be continuing to promote physical activity, even with the restrictions due to COVID-19. This finding was also supported by our qualitative results, with players consistently mentioning that AR games were keeping them motivated to exercise during the pandemic. This is particularly important because our results also show that reduced exercise is a significant predictor of poor mental health, even after taking participants' gender and age into account.

Further, consistent with emerging research [31], our study suggests that the impact of COVID-19 lockdown on mental well-being is high. Over half of participants in this study (925/1766, 52.4%) reported poor mental health. Compared to the previous estimated prevalence of poor mental health of around 25% [32], our identified prevalence of over 50% suggests that these numbers have doubled during the COVID-19 pandemic. A similar substantial impact of the current pandemic situation on mental health has been identified in other recent studies [31,33]. For example, research on large nationally representative samples of adults in the United States reported that depression symptom prevalence has more than doubled during the COVID-19 pandemic, with 52.5% of the population showing symptoms of mild depression or greater, compared to 24.7% before the pandemic [34]. The results from this emerging research on the mental health impact of COVID-19 are considerable cause for concern.

2. Impact of COVID-19–Related Social Restrictions on Motivation to Play and Video Game Use Among AR Game Players

Consistent with reports from popular media, this study found a significant increase in the use of video games during COVID-19 lockdown, mostly driven by casual gamers—women in particular—increasing their video game play. Despite the inconveniences to normal gameplay caused by COVID-19 restrictions, nearly all respondents continued to play Pokémon GO and Harry Potter: Wizards Unite. In fact, the qualitative results indicated that the in-game changes Niantic made to Pokémon GO and Harry Potter: Wizards Unite were a key motivating factor in players’ continued use of the game during the pandemic, with many reporting that they “loved” the changes or had returned to the game because of them. This is also consistent with recent research on social media reactions across three popular Pokémon GO subreddits, which showed overwhelming appreciation toward the in-game changes to Pokémon GO [10].

3. Potential Role of AR Games in Supporting the Mental Well-being of Players During COVID-19–Related Social Restrictions

Despite previous concerns about increased video game use during COVID-19 lockdown [6], more than three in four participants in this study suggested that video games helped support their mental health. The qualitative results suggest that during ongoing lockdown and social distancing measures, video games provided an escape from the fear accompanying the pandemic and provided players with something to do while their usual routines (eg, work, school, hobbies) were disrupted. Beyond that, video games were also being used to aid emotional coping—to lower stress, relax, and alleviate specific mental health conditions. Respondents also reported that games played a critical role in maintaining social connections and encouraging physical activity. Recent research has highlighted the importance of social connection [35,36] and exercise [37] to mitigate the negative psychological consequences of the COVID-19 pandemic; thus, these games represent a beneficial coping mechanism for their players.

Limitations

Our study was limited in ways that must be taken into consideration when interpreting the results. First, although the study drew participants from 66 countries, the survey was conducted in the English language only, thus excluding participation by non-English speakers. Examining cultural effects on AR game use and gamers’ well-being could be an interesting area for future research. Second, our survey was cross-sectional; therefore, we are unable to imply causality between game play and physical activity based on the current study alone. However, we have examined the data in ways that provide clear direction for future longitudinal research. Third, our study may be subject to sample bias, with recruitment occurring via four subreddit forums dedicated to Pokémon GO or Harry Potter: Wizards Unite; thus, the results may not be generalizable to the entire population. Fourth, this study also relied on self-reported data, which creates potential for recall bias. For our mental health assessment, we relied on the WHO-5 as a noninvasive assessment of subjective well-being. Although this measure cannot be used to clinically diagnose mental illness, previous research supports the use of the WHO-5 to effectively screen for depression. However, further research would benefit from the inclusion of more specific mental health issues, such as the Depression Anxiety Stress Scales [38] and the UCLA Loneliness Scale, to assess social isolation [39]. Women are also known to be less likely to self-identify as people who play video games than men [23,40,41]; this may have contributed to under-reporting of the number of participants who were “hardcore gamers,” especially for female respondents. Finally, the international scope of our sample also means that different participants had different COVID-19 lockdown experiences, as some countries enforced strict lockdowns at the time of the survey while others imposed few social limitations [42].

Comparison With Prior Work

Prior to the COVID-19 pandemic, the potential of video games to improve people’s physical and mental well-being [15] had already been well demonstrated. Beyond providing entertainment, games have the capacity to foster positive emotions [43,44], create engagement and commitment [45,46], encourage new and long-lasting social relationships [47], provide purpose and meaning to daily life [48], and deliver a sense of accomplishment and competence [46,49]. Exergames, defined as video games that mediate physical exercise [50], have been reported to be effective in clinical rehabilitation [51,52] and general exercise promotion [53,54]. They also provide holistic, cross-sectional improvements in mental health [55,56]. Within the first six months of the release of Pokémon GO, the game had already been recognized as an extremely complex behavioral intervention [57,58] that increased the duration of physical exercise for all its players, regardless of socioeconomic status [13]. Individuals and organizations worldwide also started to use games during the COVID-19 shutdown as a means of health education, mental coping, social bonding, and providing a sense of control and routine [59]. Our findings, which indicated that Pokémon GO and Harry Potter: Wizards Unite players are supporting players’ physical and mental well-being, support much of this earlier work.

A recent report from Canada [60] stated that adults are spending more time on “screens” during the pandemic. The authors of this study argued that an increase in sedentary screen time may be contributing to a decline in mental well-being during COVID-19 lockdown. In line with this argument, we found that reduced exercise was a significant predictor of poor mental health; this finding is consistent with previous pre–COVID-19 research correlating increased exercise with positive mental health [61-63]. However, the focus of the Canadian study was not on AR games such as Pokémon GO and Harry Potter: Wizards Unite. Our study suggested that playing more Pokémon GO and Harry Potter: Wizards Unite maintained participants' weekly exercise well above the recommended levels of 150 minutes of moderate exercise per week [64] and helped preserve players' mental well-being. Specifically, our study indicated that Pokémon GO and Harry Potter: Wizards Unite motivated participants to maintain a walking and socialization schedule. These behaviors reduced loneliness, which is known to correlate with worse mental health during the COVID-19 shutdown [35,65,66]. Interestingly, our qualitative findings identified that casual players, who previously played Pokémon GO and Harry Potter: Wizards Unite less frequently than the hardcore players, are now playing as much as members of the hardcore group. Further, the qualitative findings suggest that casual players, who are usually less engaged with games, are using games to both distract themselves from the pandemic and provide themselves with a sense of challenge and achievement that they usually source from activities that are currently disrupted by the pandemic.

Participants in this study indicated that they were using games in general to stabilize their mood and prevent further emotional stress. Overwhelmingly, the participants indicated that video games were beneficial to their mental health. Holistically, AR games such as Pokémon GO and Harry Potter: Wizards Unite provide virtual socialization with other people, motivate players to sustain exercise, and offer a daily routine during the lockdown, therefore establishing mental coping strategies to benefit well-being during COVID-19 lockdown. As stated recently in a commentary on gaming in the time of COVID-19, Kriz [67] suggested that "games have the potential to entertain and relax, to lower stress (maybe even provide an escape for a while from the real problems and fear accompanying the pandemic) and to provide the opportunity to interact with friends through (virtual) forms of play."

Conclusions

Location-based AR games such as Pokémon GO and Harry Potter: Wizards Unite already held the potential to be deployed as digital interventions with a beneficial impact on physical and mental health before the COVID-19 pandemic. We provide insight into the effects of continued playing of Pokémon GO or Harry Potter: Wizards Unite on a global sample of AR players. Participants reported increased game play, which coincided with their maintaining exercise levels well above international recommended levels, and using these games as a means of mental coping. Increased game play was significantly associated with increased exercise and was not associated with worse mental health during COVID-19 lockdown.

Mentally, our participants reported they continued to play the games out of a desire for entertainment, to have some structure in their daily lives, to have a feeling of accomplishment, and as a means to continue exercising. On a larger scale, the participants played video games to cope with the negative psychological impact of COVID-19 lockdown. They used games to distract themselves, escape briefly from reality, have something to do during their days, manage their emotions and mental health, and stay socially connected with other people. User goals such as these indicate that Pokémon GO and Harry Potter: Wizards Unite were already being used as veritable digital behavioral interventions by the participants to self-regulate emotional distress, especially during the periods of intensive social isolation associated with the COVID-19 pandemic.

By understanding the phenomenon of how game players worldwide are using the interactive and immersive medium of video games to manage their own mental health, we are glimpsing the massive potential video games have to become behavioral "medicines" for populations under stress. Location-based AR games have greater potential due to their propensity to encourage exercise and social connections. It would be beneficial for the health care system and digital health professionals to leverage this interactive software effectively during the mental health crisis emerging during COVID-19 lockdown, providing virtual relief and maintaining holistic well-being.

Acknowledgments

The authors would like to thank players of Pokémon GO and Harry Potter: Wizards Unite who participated in this study and the subreddit moderators for kindly assisting with survey promotion. We would like to thank and acknowledge research assistants Dr Kathryn Gibbons, Ms Sarah Hatem, and Ms Kelly Nguyen for their editorial assistance. The work was supported by the organizations involved and research grants from the National Health and Medical Research Council (APP9100002, APP1134459, APP1176620) and the Medical Research Future Fund (APP1178554).

Authors' Contributions

LAE and MDL analyzed the qualitative data; LAE and KI undertook the quantitative data analysis; and LAE took the lead in writing the manuscript. KY conceived the original idea, deployed the survey, and recruited participants. All authors were involved in the development of the web-based survey. All authors provided critical feedback and helped shape the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions asked in the web-based survey presented in this study.
[PDF File (Adobe PDF File), 105 KB - jmir_v22i12e25117_app1.pdf]

Multimedia Appendix 2
Multivariate binary logistic regression analysis.
[PDF File (Adobe PDF File), 165 KB - jmir_v22i12e25117_app2.pdf]

Multimedia Appendix 3
Summary themes identified for motivation to play Pokémon GO or Harry Potter: Wizards Unite (N=1527).
[PDF File (Adobe PDF File), 114 KB - jmir_v22i12e25117_app3.pdf]

Multimedia Appendix 4
Summary themes identified for the impact of video games on mental health (N=1427).
[PDF File (Adobe PDF File), 232 KB - jmir_v22i12e25117_app4.pdf]

References

1. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). World Health Organization. URL: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf [accessed 2020-10-15]
2. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
3. Coronavirus disease 2019 (COVID-19) Situation Report 72. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/331685/nCoVsitrep01Apr2020-eng.pdf [accessed 2020-10-15]
4. Hale T, Angrist N, Cameron-Blake E, Hallas L, Petherick A, Kira B, et al. Variation in government responses to COVID-19, Version 7. Blavatnik School of Government Working Paper Series. 2020 Sep. URL: https://www.bsg.ox.ac.uk/sites/default/files/2020-09/BSG-WP-2020-032-v7.0.pdf [accessed 2020-12-09]
5. 3, 2, 1 GO! Video Gaming is at an All-Time High During COVID-19. Nielsen. 2020 Jun 03. URL: https://www.nielsen.com/us/en/insights/article/2020/3-2-1-go-video-gaming-is-at-an-all-time-high-during-covid-19/ [accessed 2020-10-15]
6. King DL, Delfabbro PH, Billieux J, Potenza MN. Problematic online gaming and the COVID-19 pandemic. J Behav Addict 2020 Jun;9(2):184-186. [doi: 10.1556/2006.2020.00016] [Medline: 32352927]
7. #PlayApartTogether. Big Fish Games. URL: https://www.bigfishgames.com/us/en/play-apart-together.html [accessed 2020-10-15]
8. Saunders JB, Hao W, Long J, King DL, Mann K, Fauth-Bühler M, et al. Gaming disorder: Its delineation as an important condition for diagnosis, management, and prevention. J Behav Addict 2017 Sep 01;6(3):271-279 [FREE Full text] [doi: 10.1556/2006.6.2017.039] [Medline: 28816494]
9. Son JS, Nimrod G, West ST, Janke MC, Liechty T, Naar JJ. Promoting Older Adults' Physical Activity and Social Well-Being during COVID-19. Leis Sci 2020 Jun 26:1-8. [doi: 10.1080/01490400.2020.1774015]
10. Laato S, Islam AN, Laine TH. Did location-based games motivate players to socialize during COVID-19? Telemat Inform 2020 Nov;54:101458. [doi: 10.1016/j.tele.2020.101458]
11. Get Up and Go. Niantic. 2015. URL: https://nianticlabs.com/en/products [accessed 2020-10-15]
12. Zach F, Tussyadiah I. To catch them all?the (un)intended consequences of Pokémon GO on mobility, consumption, and wellbeing. In: Information and Communication Technologies in Tourism. Cham, Switzerland: Springer International; 2017:217-227.
13. Althoff T, White RW, Horvitz E. Influence of Pokémon Go on Physical Activity: Study and Implications. J Med Internet Res 2016 Dec 06;18(12):e315 [FREE Full text] [doi: 10.2196/jmir.6759] [Medline: 27923778]
14. Vella K, Johnson D, Cheng VWS, Davenport T, Mitchell J, Klarkowski M, et al. A Sense of Belonging: Pokémon GO and Social Connectedness. Games Cult 2017 Jul 20;14(6):583-603. [doi: 10.1177/1555412017719973]
15. Jones CM, Scholes L, Johnson D, Katsikitis M, Carras MC. Gaming well: links between videogames and flourishing mental health. Front Psychol 2014;5:260 [FREE Full text] [doi: 10.3389/fpsyg.2014.00260] [Medline: 24744743]
16. Laato S, Laine TH, Islam AN. Location-Based Games and the COVID-19 Pandemic: An Analysis of Responses from Game Developers and Players. MTI 2020 Jun 17;4(2):29. [doi: 10.3390/mti4020029]
17. Valentine R. Pokémon Go revenue spikes as game continues to adjust for COVID-19. gamesindustry.biz. 2020 Mar 30. URL: https://www.gamesindustry.biz/articles/2020-03-30-pokemon-go-revenue-spikes-as-game-continues-to-adjust-for-covid-19 [accessed 2020-10-15]
18. Qualtrics. 2014. URL: http://www.qualtrics.com/ [accessed 2020-10-15]

19. Ip B, Jacobs G. Segmentation of the games market using multivariate analysis. J Target Meas Anal Mark 2005 Apr 1;13(3):275-287. [doi: 10.1057/palgrave.jt.5740154]
20. Manero B, Torrente J, Freire M, Fernández-Manjón B. An instrument to build a gamer clustering framework according to gaming preferences and habits. Comput Hum Behav 2016 Sep;62:353-363. [doi: 10.1016/j.chb.2016.03.085]
21. Tuunanen J, Hamari J. Meta-synthesis of player typologies. In: Proceedings of 2012 International Digra Nordic Conference. 2012 Presented at: 2012 International Digra Nordic Conference; June 6-8, 2012; Tampere, Finland URL: http://www.digra.org/wp-content/uploads/digital-library/12168.40312.pdf
22. Kuittinen J, Kultima A, Niemelä J, Paavilainen J. Casual games discussion. In: Proceedings of the 2007 Conference on Future Play. 2007 Nov Presented at: Future Play '07; November 2007; Toronto, ON p. 105-112. [doi: 10.1145/1328202.1328221]
23. Yee N. What Men and Women Consider Hardcore Gaming Are Not The Same. Quantic Foundry. 2018 Aug 01. URL: https://quanticfoundry.com/2018/08/01/casual-hardcore/#:~:text=all%20gaming%20motivations.-,Female%20Hardcore%20gamers%20are%20consistently%20above%20average%20across%20all%20gaming,in%20all%20aspects%20of%20gaming [accessed 2020-12-09]
24. The WHO-5 website. URL: https://www.psykiatri-regionh.dk/who-5/Pages/default.aspx [accessed 2020-10-15]
25. Halliday JA, Hendrieckx C, Busija L, Browne JL, Nefs G, Pouwer F, et al. Validation of the WHO-5 as a first-step screening instrument for depression in adults with diabetes: Results from Diabetes MILES - Australia. Diabetes Res Clin Pract 2017 Oct;132:27-35. [doi: 10.1016/j.diabres.2017.07.005] [Medline: 28783530]
26. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
27. NVivo qualitative data analysis software. QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2020-12-09]
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
29. Giorgi A. The Theory, Practice, and Evaluation of the Phenomenological Method as a Qualitative Research Procedure. J Phenomenol Psychol 1997;28(2):235-260. [doi: 10.1163/156916297x00103]
30. Global Recommendations on Physical Activity for Health. Geneva, Switzerland: World Health Organization; 2010.
31. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
32. Dreger S, Buck C, Bolte G. Material, psychosocial and sociodemographic determinants are associated with positive mental health in Europe: a cross-sectional study. BMJ Open 2014 May 28;4(5):e005095 [FREE Full text] [doi: 10.1136/bmjopen-2014-005095] [Medline: 24871540]
33. Bueno-Notivol J, Gracia-García P, Olaya B, Lasheras I, López-Antón R, Santabárbara J. Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies. Forthcoming 2020. Int J Clin Health Psychol [FREE Full text] [doi: 10.1016/j.ijchp.2020.07.007] [Medline: 32904715]
34. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
35. Tull MT, Edmonds KA, Scamaldo KM, Richmond JR, Rose JP, Gratz KL. Psychological Outcomes Associated with Stay-at-Home Orders and the Perceived Impact of COVID-19 on Daily Life. Psychiatry Res 2020 May 12;289:113098 [FREE Full text] [doi: 10.1016/j.psychres.2020.113098] [Medline: 32434092]
36. Marston H, Musselwhite C, Hadley R. COVID-19 vs Social Isolation: the impact technology can have on communities, social connections and citizens. The British Society of Gerontology. 2020 Mar 18. URL: https://ageingissues.wordpress.com/2020/03/18/covid-19-vs-social-isolation-the-impact-technology-can-have-on-communities-social-connections-and-citizens/ [accessed 2020-12-09]
37. Hammami A, Harrabi B, Mohr M, Krustrup P. Physical activity and coronavirus disease 2019 (COVID-19): specific recommendations for home-based physical training. Manag Sport Leis 2020 Apr 20:1-6. [doi: 10.1080/23750472.2020.1757494]
38. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. 2nd edition. Sydney, Australia: Psychology Foundation of Australia; 1995.
39. Hays R, DiMatteo MR. A short-form measure of loneliness. J Pers Assess 1987 Mar 1;51(1):69-81. [doi: 10.1207/s15327752jpa5101_6] [Medline: 3572711]
40. Duggan M. Gaming and Gamers. Pew Research Center. 2015 Dec 15. URL: https://www.pewresearch.org/internet/2015/12/15/gaming-and-gamers/ [accessed 2020-12-09]
41. Shaw A. Do you identify as a gamer? Gender, race, sexuality, and gamer identity. New Media Soc 2011 Jun 16;14(1):28-44. [doi: 10.1177/1461444811410394]
42. Braithwaite J, Tran Y, Ellis L, Westbrook J. The 40 health systems, COVID-19 (40HS, C-19) study. Int J Qual Health Care 2020 Sep 30:mzaa113 [FREE Full text] [doi: 10.1093/intqhc/mzaa113] [Medline: 32991713]

43. Kutner L, Olson C. Grand Theft Childhood: The Surprising Truth About Violent Video Games and What Parents Can Do. New York, NY: Simon and Schuster; 2008.
44. Przybylski AK, Weinstein N, Murayama K, Lynch MF, Ryan RM. The ideal self at play: the appeal of video games that let you be all you can be. Psychol Sci 2012 Jan 01;23(1):69-76. [doi: 10.1177/0956797611418676] [Medline: 22173739]
45. Snodgrass JG, Lacy MG, Francois Dengah HJ, Fagan J, Most DE. Magical flight and monstrous stress: technologies of absorption and mental wellness in Azeroth. Cult Med Psychiatry 2011 Mar;35(1):26-62. [doi: 10.1007/s11013-010-9197-4] [Medline: 21165683]
46. Snodgrass JG, Lacy MG, Francois Dengah H, Fagan J. Enhancing one life rather than living two: Playing MMOs with offline friends. Computers in Human Behavior 2011 May;27(3):1211-1222. [doi: 10.1016/j.chb.2011.01.001]
47. Cole H, Griffiths MD. Social interactions in massively multiplayer online role-playing gamers. Cyberpsychol Behav 2007 Aug;10(4):575-583. [doi: 10.1089/cpb.2007.9988] [Medline: 17711367]
48. McGonigal J. Reality Is Broken: Why Games Make Us Better and How They Can Change the World. Westminster, UK: Penguin Books; Dec 07, 2011.
49. Ryan RM, Rigby CS, Przybylski A. The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
50. Oh Y, Yang S. 2010 Presented at: Meaningful Play; October 21-23, 2010; East Lansing, MI URL: https://meaningfulplay.msu.edu/proceedings2010/mp2010_paper_63.pdf
51. Şimşek TT, Çekok K. The effects of Nintendo Wii(TM)-based balance and upper extremity training on activities of daily living and quality of life in patients with sub-acute stroke: a randomized controlled study. Int J Neurosci 2016 Dec;126(12):1061-1070. [doi: 10.3109/00207454.2015.1115993] [Medline: 26626539]
52. Taylor MJD, Griffin M. The use of gaming technology for rehabilitation in people with multiple sclerosis. Mult Scler 2015 Apr;21(4):355-371. [doi: 10.1177/1352458514563593] [Medline: 25533296]
53. Chao Y, Scherer YK, Montgomery CA, Wu Y, Lucke KT. Physical and psychosocial effects of Wii Fit exergames use in assisted living residents: a pilot study. Clin Nurs Res 2015 Dec;24(6):589-603. [doi: 10.1177/1054773814562880] [Medline: 25488422]
54. Peng W, Crouse JC, Lin J. Using active video games for physical activity promotion: a systematic review of the current state of research. Health Educ Behav 2013 Apr;40(2):171-192. [doi: 10.1177/1090198112444956] [Medline: 22773597]
55. Li J, Theng Y, Foo S. Effect of Exergames on Depression: A Systematic Review and Meta-Analysis. Cyberpsychol Behav Soc Netw 2016 Jan;19(1):34-42. [doi: 10.1089/cyber.2015.0366] [Medline: 26716638]
56. Rosenberg D, Depp CA, Vahia IV, Reichstadt J, Palmer BW, Kerr J, et al. Exergames for subsyndromal depression in older adults: a pilot study of a novel intervention. Am J Geriatr Psychiatry 2010 Mar;18(3):221-226 [FREE Full text] [doi: 10.1097/JGP.0b013e3181c534b5] [Medline: 20173423]
57. Clark A, Clark M. Pokémon Go and Research. Int J Qual Methods 2016 Aug 29;15(1):160940691666776. [doi: 10.1177/1609406916667765]
58. Tateno M, Skokauskas N, Kato TA, Teo AR, Guerrero APS. New game software (Pokémon Go) may help youth with severe social withdrawal, hikikomori. Psychiatry Res 2016 Dec 30;246:848-849 [FREE Full text] [doi: 10.1016/j.psychres.2016.10.038] [Medline: 27817905]
59. Balakrishnan VS. COVID-19: Playing away the pandemic. Lancet Infect Dis 2020 Jul;20(7):792 [FREE Full text] [doi: 10.1016/S1473-3099(20)30507-7] [Medline: 32592678]
60. Colley R, Bushnik T, Langlois K. Exercise and screen time during the COVID-19 pandemic. Health Rep 2020 Jul 15;31(6):3-11 [FREE Full text] [doi: 10.25318/82-003-x202000600001-eng] [Medline: 32672923]
61. Taylor CB, Sallis JF, Needle R. The relation of physical activity and exercise to mental health. Public Health Rep 1985;100(2):195-202 [FREE Full text] [Medline: 3920718]
62. Mikkelsen K, Stojanovska L, Polenakovic M, Bosevski M, Apostolopoulos V. Exercise and mental health. Maturitas 2017 Dec;106:48-56. [doi: 10.1016/j.maturitas.2017.09.003] [Medline: 29150166]
63. Chekroud SR, Gueorguieva R, Zheutlin AB, Paulus M, Krumholz HM, Krystal JH, et al. Association between physical exercise and mental health in 1·2 million individuals in the USA between 2011 and 2015: a cross-sectional study. Lancet Psychiatry 2018 Sep;5(9):739-746. [doi: 10.1016/S2215-0366(18)30227-X] [Medline: 30099000]
64. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The Physical Activity Guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
65. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
66. Liu CH, Zhang E, Wong GTF, Hyun S, Hahm HC. Factors associated with depression, anxiety, and PTSD symptomatology during the COVID-19 pandemic: Clinical implications for U.S. young adult mental health. Psychiatry Res 2020 Aug;290:113172 [FREE Full text] [doi: 10.1016/j.psychres.2020.113172] [Medline: 32512357]
67. Kriz WC. Gaming in the Time of COVID-19. Simulation & Gaming 2020 Jul 14;51(4):403-410. [doi: 10.1177/1046878120931602]

Abbreviations

ANOVA: analysis of variance
AR: augmented reality
WHO: World Health Organization
WHO-5: World Health Organization–5 Well-Being Index

Edited by G Eysenbach; submitted 18.10.20; peer-reviewed by J Li, H Söbke; comments to author 09.11.20; revised version received 24.11.20; accepted 05.12.20; published 22.12.20.

Please cite as:
Ellis LA, Lee MD, Ijaz K, Smith J, Braithwaite J, Yin K
COVID-19 as 'Game Changer' for the Physical Activity and Mental Well-Being of Augmented Reality Game Players During the Pandemic: Mixed Methods Survey Study
J Med Internet Res 2020;22(12):e25117
URL: http://www.jmir.org/2020/12/e25117/
doi:10.2196/25117
PMID:33284781

Original Paper

Digital Monitoring and Management of Patients With Advanced or Metastatic Non-Small Cell Lung Cancer Treated With Cancer Immunotherapy and Its Impact on Quality of Clinical Care: Interview and Survey Study Among Health Care Professionals and Patients

Oliver Schmalz[1*], MD; Christine Jacob[2,3*], MSc, PhD; Johannes Ammann[4*], MSc, PhD; Blasius Liss[1], MD; Sanna Iivanainen[5], MD, PhD; Manuel Kammermann[6], MSc; Jussi Koivunen[5], MD, PhD; Alexander Klein[4], MSc; Razvan Andrei Popescu[7], MD

[1]Department of Haematology, Oncology and Palliative Care, Helios Universitätsklinikum Wuppertal, Wuppertal, Germany
[2]The Faculty of Business and Law, Anglia Ruskin University, Cambridge, United Kingdom
[3]School of Business, University of Applied Sciences Northwestern Switzerland, Brugg, Switzerland
[4]Medical Affairs (Personalised Healthcare and Patient Access), F Hoffmann-La Roche Ltd, Basel, Switzerland
[5]Department of Oncology and Radiotherapy, Medical Research Center Oulu, Oulu University Hospital, Oulu, Finland
[6]Arcondis AG, Reinach, Switzerland
[7]Tumor Zentrum Aarau, Hirslanden Klinik Aarau, Aarau, Switzerland
[*]these authors contributed equally

Corresponding Author:
Johannes Ammann, MSc, PhD
Medical Affairs (Personalised Healthcare and Patient Access)
F Hoffmann-La Roche Ltd
Grenzacherstrasse 124
Basel, 4070 Basel
Switzerland
Phone: 41 792 921 758
Email: johannes.ammann@roche.com

Abstract

Background: Cancer immunotherapy (CIT), as a monotherapy or in combination with chemotherapy, has been shown to extend overall survival in patients with locally advanced or metastatic non-small cell lung cancer (NSCLC). However, patients experience treatment-related symptoms that they are required to recall between hospital visits. Digital patient monitoring and management (DPMM) tools may improve clinical practice by allowing real-time symptom reporting.

Objective: This proof-of-concept pilot study assessed patient and health care professional (HCP) adoption of our DPMM tool, which was designed specifically for patients with advanced or metastatic NSCLC treated with CIT, and the tool's impact on clinical care.

Methods: Four advisory boards were assembled in order to co-develop a drug- and indication-specific CIT (CIT+) module, based on a generic CIT DPMM tool from Kaiku Health, Helsinki, Finland. A total of 45 patients treated with second-line single-agent CIT (ie, atezolizumab or otherwise) for advanced or metastatic NSCLC, as well as HCPs, whose exact number was decided by the clinics, were recruited from 10 clinics in Germany, Finland, and Switzerland between February and May 2019. All clinics were provided with the Kaiku Health generic CIT DPMM tool, including our CIT+ module. Data on user experience, overall satisfaction, and impact of the tool on clinical practice were collected using anonymized surveys—answers ranged from 1 (low agreement) to 5 (high agreement)—and HCP interviews; surveys and interviews consisted of closed-ended Likert scales and open-ended questions, respectively. The first survey was conducted after 2 months of DPMM use, and a second survey and HCP interviews were conducted at study end (ie, after ≥3 months of DPMM use); only a subgroup of HCPs from each clinic responded to the surveys and interviews. Survey data were analyzed quantitatively; interviews were recorded, transcribed verbatim, and translated into English, where applicable, for coding and qualitative thematic analysis.

Results: Among interim survey respondents (N=51: 13 [25%] nurses, 11 [22%] physicians, and 27 [53%] patients), mean rankings of the tool's seven usability attributes ranged from 3.2 to 4.4 (nurses), 3.7 to 4.5 (physicians), and 3.7 to 4.2 (patients). At the end-of-study survey (N=48: 19 [40%] nurses, 8 [17%] physicians, and 21 [44%] patients), most respondents agreed that the tool facilitated more efficient and focused discussions between patients and HCPs (nurses and patients: mean rating 4.2, SD 0.8; physicians: mean rating 4.4, SD 0.8) and allowed HCPs to tailor discussions with patients (mean rating 4.35, SD 0.65). The standalone tool was well integrated into HCP daily clinical workflow (mean rating 3.80, SD 0.75), enabled workflow optimization between physicians and nurses (mean rating 3.75, SD 0.80), and saved time by decreasing phone consultations (mean rating 3.75, SD 1.00) and patient visits (mean rating 3.45, SD 1.20). Workload was the most common challenge of tool use among respondents (12/19, 63%).

Conclusions: Our results demonstrate high user satisfaction and acceptance of DPMM tools by HCPs and patients, and highlight the improvements to clinical care in patients with advanced or metastatic NSCLC treated with CIT monotherapy. However, further integration of the tool into the clinical information technology data flow is required. Future studies or registries using our DPMM tool may provide insights into significant effects on patient quality of life or health-economic benefits.

(J Med Internet Res 2020;22(12):e18655) doi:10.2196/18655

KEYWORDS

advanced or metastatic non-small cell lung cancer; cancer immunotherapy; digital patient monitoring; drug- and indication-specific cancer immunotherapy module; eHealth; mHealth; quality of patient care; patient-reported outcomes; real-time symptom reporting; user experience

Introduction

Lung cancer was the most common, newly diagnosed malignancy and leading cause of cancer-related death worldwide in 2018 [1]; approximately 85% of cases are classified as non-small cell lung cancer (NSCLC) [2]. Guidelines for patients with locally advanced or metastatic NSCLC without alterations in *EGFR* (epidermal growth factor receptor), *ALK* (anaplastic lymphoma kinase), *ROS1* (ROS proto-oncogene 1), *BRAF* (B-Raf proto-oncogene), *NTRK* (neurotrophic tropomyosin receptor kinase), *RET* (rearranged during transfection), or *MET* (N-methyl-N'-nitroso-guanidine human osteosarcoma transforming) genes, for which targeted therapies are available, recommend first-line treatment with the cancer immunotherapy (CIT) pembrolizumab, as monotherapy for patients with 50% or higher tumor cell programmed death-ligand 1 (PD-L1)–positivity or in combination with chemotherapy [3,4]. Atezolizumab plus bevacizumab, paclitaxel, and carboplatin is also indicated for first-line treatment of patients with metastatic NSCLC and any PD-L1 expression level [5]. For patients who have not previously received CIT treatment, second-line nivolumab, pembrolizumab, or atezolizumab may be given following first-line chemotherapy [3,4].

CIT regimens activate the immune system against cancer and have been shown to slow disease progression and extend overall survival (OS) versus standard chemotherapy and when added to standard chemotherapy [6-13]. However, many patients treated with CIT experience related side effects, such as fatigue, skin rash or itching, diarrhea, nausea, vomiting, dyspnea, and cough [14], in addition to NSCLC-related symptoms [15]. These symptoms can be identified and reported by patients during clinic visits as per the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE). This patient-reported outcome (PRO) collection has been shown to improve patient-clinician communication, improve patient satisfaction, and enable early symptom detection [16-18]. This has led to the development of PRO-CTCAE, a measurement system designed by the NCI as a companion to the CTCAE to evaluate symptom toxicity [19]. However, the need to rely on patients to recall symptom type and severity over a certain time period between visits can lead to health care professionals (HCPs) receiving incomplete information, thus preventing efficient management.

Digital patient monitoring and management (DPMM) tools may improve clinical practice by allowing patients to report symptoms in real time, enabling direct patient-HCP communication and providing access to patient support materials [20]. However, in addition to collecting and aggregating symptom information weekly, they can also improve patient OS and quality of life (QoL), as well as offer health-economic benefits, such as reduced hospital admission rates and unscheduled visits [21-23]. A study of patients with advanced nonprogressive stage IIA-IV lung cancer finishing first-line chemotherapy found a significantly improved median OS with web-based symptom monitoring versus standard scheduled imaging after a 2-year follow-up: 22.5 versus 14.9 months (hazard ratio 0.59, 95% CI 0.37-0.96, *P*=.03) [24]. To achieve these benefits, DPMM tools must be adopted easily into clinical practice and used frequently so that critical symptoms are reported and detected, and care initiated as early as possible, which is particularly important for increasing OS [25].

In this proof-of-concept pilot study, we assessed factors influencing patient and HCP adoption of our DPMM tool, designed and developed specifically for patients with CIT-treated advanced or metastatic NSCLC, and the impact of such adoption on the quality of clinical care. The study focused on patients treated with CIT due to the high unmet medical need for early detection of critical symptoms in this subgroup. Our tool was based on the generic CIT DPMM tool developed by Kaiku Health in Helsinki, Finland, including all basic features, as well as additional drug-specific features. The Kaiku Health platform was selected due to Kaiku Health's focus on oncology, including CIT; its market availability in five European Union countries and Switzerland; and its established use in routine practice.

Methods

Recruitment

HCP and patient participants were recruited from 10 clinics across Germany, Switzerland, and Finland between February and May 2019. This study used purposive sampling; potential participants with tool experience were selected so that they could provide in-depth information about the research topic [26,27]. Roche developed paper-based materials to support oncologists with patient recruitment, including a welcome letter and device-specific instructions for the platform. Of the 10 clinics involved in the pilot study, three were already using Kaiku Health's generic CIT DPMM tool and seven had only limited experience with other DPMM platforms. Participating clinics ranged from small community clinics to large university hospitals to reflect the natural diversity of cancer care centers. A single point of contact within each clinic decided on the exact number of HCP participants. A total of 56 patients with advanced or metastatic NSCLC treated with second-line single-agent CIT (ie, atezolizumab or otherwise) were recruited. The final number of included patients was 45, 9 (20%) of whom were treated with atezolizumab; out of 56 patients originally recruited, 5 (9%) declined the invitation and 6 (11%) withdrew early due to disease progression.

Developing a Drug- and Indication-Specific CIT Module

A literature review was conducted to define CIT-related symptoms and to identify key factors influencing DPMM tool use. Four separate advisory boards were also assembled, with meetings conducted in November 2018 to explore expectations, perceived value, and concerns of HCPs and patients with regard to DPMM tools; these included two boards for physicians (oncologists; n=4), one for nurses (n=4), and one for patients (n=1). The information obtained was used to co-develop a drug- and indication-specific CIT (CIT+) module for Kaiku Health's DPMM platform, centering on patients' and HCPs' needs (see Multimedia Appendix 1). Kaiku Health, as the technical partner, provided the existing, generic CIT DPMM tool, which was used to co-develop the CIT+ module. Kaiku Health's generic CIT DPMM tool, including the CIT+ module, was provided to the 10 participating clinics, being made available on smartphones, tablets, and desktop computers. Patients treated with a CIT other than atezolizumab had access to functions such as a symptom questionnaire, as per PRO-CTCAE, with 18 questions specific for NSCLC CIT monotherapy; direct message communication between patients and HCPs; indication-specific educational material with information on mild to moderate symptoms and their management [28,29]; and a symptom overview and alerts for HCPs. Patients treated with atezolizumab had access to the above functions and additional drug-specific educational material (eg, a patient card and information on preparing for first infusion and treatment and the likelihood of symptom incidence). Symptom reporting within the tool was required weekly as per the NCI PRO-CTCAE guidelines and based on recommendations from HCPs and patients and usage frequency in seminal clinical trials [22,30]. On-site onboarding sessions of 2 hours in length were held at each clinic between February and April 2019 to train care teams on the CIT+ module and to address questions (see Multimedia Appendix 2). Training included overviews of the pilot study, the partner Kaiku Health, the platform, and triaging workflow. Practical role-based training with simulations, including patient onboarding, electronic PRO (ePRO) form-filling, triaging results, and care team–patient communication, were provided.

Data Collection

To test the DPMM tool in a real-world setting, data on user experience, overall satisfaction, and the impact of the tool on clinical practice were collected using anonymized surveys and HCP interviews. Fidelity of delivery of the tool and patient adherence were measured using self-reported data in the survey and system usage statistics obtained from the tool itself. Patient surveys were provided to patients by HCPs. An online interim survey consisting of a short questionnaire with 11 closed-ended multiple-choice or Likert scale questions in English, Finnish, or German was conducted after 2 months of tool use to assess user satisfaction and to allow early identification of potential issues. At study end (ie, ≥3 months of tool use), a second online survey consisting of a long questionnaire with 34 and 36 closed-ended multiple-choice or Likert scale questions for patients and HCPs, respectively, was conducted to assess value and to highlight potential gaps or need for improvement. Surveys were built using SurveyMonkey. Semistructured interviews with HCPs in English, Finnish, or German were also conducted at study end to answer 14 open-ended questions to better understand their views on the tool and to increase understanding of survey results. Only a subset of HCPs and patients in each clinic responded to the surveys and interviews. The questions in each were informed by factors included in the original technology acceptance model (TAM) [31], notably perceived usefulness and perceived ease of use [32]. The TAM was selected for its simplicity and for being one of the most commonly used frameworks for assessing user acceptance of new technologies in general health care [33-36] and, specifically, mobile health [37]. The questions were also informed by other elements frequently reported in the literature as influencing adoption of mobile health solutions, for example, patient-clinician communications [38-41], quality of care [42-46], empowerment of patients and care teams [47-53], and efficiency [54-57]. The surveys and interview guides were tested and piloted before their use in the study.

Data Analysis

Due to the small number of included patients with advanced or metastatic NSCLC treated with atezolizumab (9/45, 20%), data for the whole CIT+ module, including both the generic CIT and the atezolizumab-specific components, were pooled. Survey data were aggregated and analyzed quantitatively using Microsoft Excel for Mac 2011, version 14.4.3 (Microsoft Corporation), to calculate totals, percentages, means, and standard deviations. HCP interviews were recorded, transcribed verbatim, and translated into English, where applicable, for coding. NVivo (QSR International) version 12.6.0 (3841), a qualitative data analysis software package, was used for coding and categorization of transcripts. Data were systematically analyzed using thematic analysis methodology; after the initial analysis and coding by CJ, this was reviewed by MK, and any cases of disagreement were discussed in conjunction with AK and mutually agreed upon (see Multimedia Appendix 3) [58,59]. Anonymized data regarding in-module activities of all included patients, including time to complete the symptoms questionnaire, use of the chat function, and engagement with educational material, were collected by Kaiku Health and shared with Roche through Chartio, software that allows for multiple individuals to access and modify data from different sources. Chartio was used to create detailed usage reports on deidentified user interactions, including log-in events, article reading times, and downloads. The report data were visualized using Data Studio from G Suite Business Solutions (Google).

Ethical Considerations

Due to the user experience nature of the study, ethics committee approval was not required for the participating sites; however, some sites submitted the study to the ethical committee on a voluntary basis and received approval. Data were anonymized, and no internet protocol data were collected. All participants gave written informed consent. Patients were contacted by their own care team only, and all treatment-related decisions were made solely by the treating physician.

Results

User Acceptance of the DPMM Tool and Overall Satisfaction

All user groups, particularly HCPs, showed an increased preference for the desktop version of the tool (see Multimedia Appendix 4). A total of 51 respondents—13 (25%) nurses, 11 (22%) physicians, and 27 (53%) patients—completed the interim survey. Respondents were asked to rank the usability attributes of the tool, with answers rated from 1 (low agreement) to 5 (high agreement). All attributes were ranked quite highly, with mean rank scores ranging from 3.2 to 4.4 for nurses, 3.7 to 4.5 for physicians, and 3.7 to 4.2 for patients (see Table 1). Across all user groups, the highest-ranking attributes were usefulness and communication, followed by ease of use, the value of onboarding, improved quality of care, empowerment, and efficiency (see Table 1). Efficiency was ranked lowest by nurses and physicians, and second lowest by patients; however, overall user acceptance was high and there was a high level of satisfaction with the tool across all user groups. For most user groups, both experienced and new, SDs were less than 1, indicating general alignment. Slight disagreement occurred for new physician users regarding usefulness of the tool (SD 1.2), communication (SD 1.0), and quality of care (SD 1.0).

Table 1. Digital patient monitoring and management tool user satisfaction among interim survey respondents.

Usability attribute	Rating[a], mean (SD)		
	Nurses (n=13)	Physicians (n=11)	Patients (n=27)
Onboarding	4.0 (0.5)	4.1 (1.0)	4.1 (0.3)
Ease of use	3.8 (0.4)	4.4 (0.4)	4.1 (0.4)
Usefulness	4.4 (0.2)	4.5 (0.4)	4.1 (0.1)
Communication	4.4 (0.5)	4.4 (0.3)	4.2 (0.3)
Efficiency	3.2 (0.2)	3.7 (0.3)	3.7 (0.4)
Empowerment	3.8 (0.2)	4.4 (0.4)	3.5 (0.3)
Quality of care	4.0 (0.5)	4.1 (0.6)	3.9 (0.5)

[a] User satisfaction of usability attributes was rated on a scale of 1 (low agreement) to 5 (high agreement).

User Statistics of End-of-Study Survey Respondents

There were 48 respondents of the end-of-study survey: 19 (40%) nurses, 8 (17%) physicians, and 21 (44%) patients. Characteristics and information on tool usage are provided in Figure 1. Most respondents were female, primarily due to the higher number of participating female nurses (see Figure 1, A). Most respondents were 40 to 70 years old and had no previous experience of using Kaiku Health or other DPMM tools before their involvement in this pilot study (see Figure 1, B and C). Overall, 35 out of 48 (73%) end-of-study survey respondents considered their proficiency level during tool use to be competent, proficient, or expert (see Figure 1, D). Frequency of tool use was at least weekly for 41 out of 48 (85%) respondents (see Figure 1, E), with 29 out of 48 (60%) respondents indicating that they used the tool for 10 minutes or less per session (see Figure 1, F).

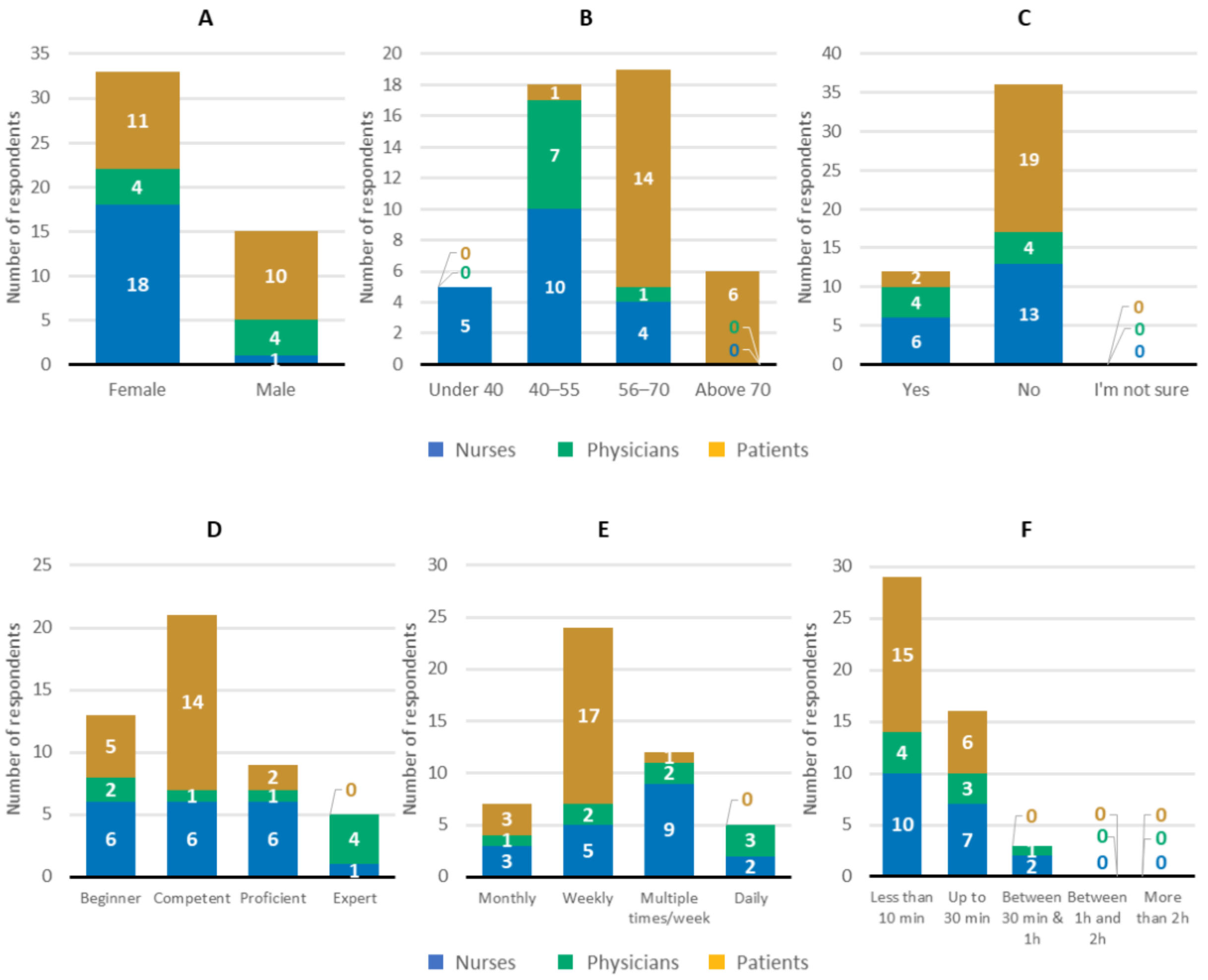


Figure 1. Characteristics of end-of-study survey respondents and digital patient monitoring and management (DPMM) tool usage (N=48: 19 [40%] nurses, 8 [17%] physicians, and 21 [44%] patients). (A) gender, (B) age, (C) DPMM use (ie, response to "Have you used Kaiku Health or other similar digital monitoring tools before this pilot?"), (D) level of proficiency (ie, response to "How would you rate your current proficiency level with regard to Kaiku Health?"), (E) usage rate (ie, response to "How often do you use Kaiku Health?"), and (F) time spent per session using the tool for clinicians and patients. Data are the number of respondents who provided the given response.

Effect of the Tool on Communication, Quality of Patient Care, and Efficiency

End-of-study survey respondents were questioned about their perceptions of the effect of the tool on communication, quality of patient care, and efficiency. All user groups agreed that the tool facilitated more efficient and focused discussions between patients and HCPs: mean ratings ranged from 4.2 for nurses (SD 0.8) and patients (SD 1.0) to 4.4 for physicians (SD 0.8) (see Table 2).

Table 2. Effect of the digital patient monitoring and management tool on patient–health care professional communication for end-of-study survey respondents.

Survey statement	Rating[a], mean (SD)		
	Nurses (n=19)	Physicians (n=8)	Patients (n=21)
Kaiku Health allows for more efficient communication with patients	4.2 (0.8)	4.4 (0.8)	N/A[b]
Kaiku Health helps to focus my discussions with my care team	N/A	N/A	4.3 (1.0)
Kaiku Health makes it easier to communicate with my care team	N/A	N/A	4.2 (1.0)

[a] Responses to survey statements were given on a scale of 1 (low agreement) to 5 (high agreement).

[b] N/A: not applicable. This user group was not presented with this statement.

Patients also indicated that the tool made communication with their care team easier: mean rating 4.3 (SD 1.0) (see Table 2). In that regard, the tool included a chat function that allowed for messages to be sent between patients and HCPs; overall, German or Finnish HCPs sent more messages to patients (n=326) than patients sent to German or Finnish HCPs (n=265), whereas the numbers of messages sent between Swiss HCPs and patients were similar: 47 versus 54 (see Multimedia Appendix 5).

Ratings from HCPs at the end-of-study survey showed that they believed that the tool helped to improve quality of patient care (mean rating 4.10, SD 0.85), permitting tailored discussions with patients (mean rating 4.35, SD 0.65), and that the symptom alert feature allowed earlier detection of symptoms (mean rating 4.25, SD 0.85) and tailoring of treatment plans (mean rating 3.9, SD 1.0). The self-care instructions function was appreciated by both HCPs and patients (mean ratings ranged from 4.0, SD 1.0, for physicians to 4.1, SD 0.9, for nurses and 4.1, SD 0.5, for patients). Patients also agreed that the tool made them feel more taken care of (mean rating 3.9, SD 1.3) (see Table 3).

Table 3. Effect of the digital patient monitoring and management tool on quality of patient care for end-of-study survey respondents.

Survey statement	Rating[a], mean (SD)		
	Nurses (n=19)	Physicians (n=8)	Patients (n=21)
Kaiku Health helps me to improve quality of patient care	3.9 (1.0)	4.3 (0.7)	N/A[b]
Kaiku Health helps me tailor my discussions with my patients	4.2 (0.8)	4.5 (0.5)	N/A
The symptom alert feature alerts my staff to react to symptoms earlier	3.9 (1.2)	4.6 (0.5)	N/A
The symptom alert feature enables my staff to tailor treatment plans	3.8 (1.0)	4.0 (1.0)	N/A
Self-care instructions are valuable	4.1 (0.9)	4.0 (1.0)	N/A
Self-care instructions make me feel informed	N/A	N/A	4.1 (0.5)
Kaiku Health makes me feel more taken care of	N/A	N/A	3.9 (1.3)

[a] Responses to survey statements were given on a scale of 1 (low agreement) to 5 (high agreement).

[b] N/A: not applicable. This user group was not presented with this statement.

HCPs, particularly physicians, thought that the standalone tool was well integrated into their daily clinical workflow (mean rating 3.80, SD 0.75). They thought that it could help to improve efficiency by enabling workflow optimization between physicians and nurses (mean rating 3.75, SD 0.80) and freeing up time by decreasing the need for phone consultations (mean rating 3.75, SD 1.00) and patient visits (mean rating 3.45, SD 1.20) through online symptom assessment (see Table 4). Patients also thought that the tool improved efficiency by improving their ability to evaluate whether their symptoms required an unscheduled outpatient appointment (mean rating 3.9, SD 1.2) through prompts to contact HCPs regarding severe symptoms. Patients further reported a shortening of the time between health consultation requests and responses (mean rating 3.7, SD 1.2) (see Table 4).

Table 4. Effect of the digital patient monitoring and management tool on efficiency for end-of-study survey respondents: general comments.

Survey statement	Rating[a], mean (SD)		
	Nurses (n=19)	Physicians (n=8)	Patients (n=21)
Kaiku Health is well integrated into my daily workflow	3.6 (1.0)	4.0 (0.5)	N/A[b]
The workflow management function between nurses and oncologists enables workflow optimization	3.4 (0.9)	4.1 (0.7)	N/A
Kaiku Health potentially decreases unnecessary patient visits and frees up time	3.4 (1.0)	3.5 (1.4)	N/A
Kaiku Health potentially decreases unnecessary patient phone calls and frees up time	3.4 (1.0)	4.1 (1.0)	N/A
Kaiku Health shortens the time between my health consultation requests and response	N/A	N/A	3.7 (1.2)
Kaiku Health helps me to better evaluate if my symptoms require a hospital visit	N/A	N/A	3.9 (1.2)

[a] Responses to survey statements were given on a scale of 1 (low agreement) to 5 (high agreement).

[b] N/A: not applicable. This user group was not presented with this statement.

For HCPs, the tool required little time for patient introduction, with most (18/27, 67%) taking up to 30 minutes for onboarding per patient. The tool also saved them time during patient visits (6/27, 22%, saved ≤5 minutes per consultation; 5/27, 19%, saved 6-10 minutes; and 1/27, 4%, saved 11-15 minutes; see Table 5) . Out of 21 patients, 3 (14%) reported that their need for an unscheduled, symptom-related hospital visit decreased per month during their use of the tool, while 1 patient (5%) reported an increased number of monthly visits and 8 (38%) reported no change in the frequency of unscheduled hospital visits. Out of 21 patients, 7 (33%) reported a decreased need for a phone consultation while using the tool; for 9 patients (43%), the need stayed the same (see Table 5).

Table 5. Effect of the digital patient monitoring and management tool on efficiency for end-of-study survey respondents: health care professional time invested or saved, and patient need for unscheduled hospital visits and telephone consultations.

Survey question or statement and responses	Number of respondents who provided the given response, n (%)[a]		
	Nurses (n=19)	Physicians (n=8)	Patients (n=21)
How long does it take to onboard the patient?			
Up to 30 minutes	13 (68)	5 (63)	N/A[b]
Between 30 minutes and 1 hour	2 (11)	0 (0)	N/A
Between 1 and 2 hours	0 (0)	0 (0)	N/A
More than 2 hours	0 (0)	0 (0)	N/A
I don’t onboard patients	4 (21)	3 (38)	N/A
Kaiku Health allows me to save time, which amounts to approximately...			
Up to 5 minutes per consultation	3 (16)	3 (38)	N/A
Between 6 and 10 minutes per consultation	4 (21)	1 (13)	N/A
Between 11 and 15 minutes per consultation	0 (0)	1 (13)	N/A
More than 16 minutes per consultation	0 (0)	0 (0)	N/A
It does not save any time	5 (26)	2 (25)	N/A
Kaiku Health needs even more time	1 (5)	0 (0)	N/A
I am not sure	6 (32)	1 (13)	N/A
Since I started using Kaiku Health, the number of unscheduled hospital visits due to observed symptoms...			
Decreased on average by 1 visit per month	N/A	N/A	1 (5)
Decreased on average by 2 visits per month	N/A	N/A	2 (10)
Decreased on average by 3 or more visits	N/A	N/A	0 (0)
Increased on average by 1 visit per month	N/A	N/A	1 (5)
Increased on average by 2 visits per month	N/A	N/A	0 (0)
Increase on average by 3 or more visits	N/A	N/A	0 (0)
Amount of visits did not change	N/A	N/A	8 (38)
I don’t know or not applicable since I started	N/A	N/A	9 (43)
Due to the use of Kaiku Health, my need to request a telephone consultation...			
Increased	N/A	N/A	0 (0)
Decreased	N/A	N/A	7 (33)
Stayed the same	N/A	N/A	9 (43)
Does not apply	N/A	N/A	5 (24)

[a] Percentages may not add up to 100% due to rounding.

[b] N/A: not applicable. This user group was not presented with this question or statement.

Exploration of HCP Needs, Expectations, and Perceived Value of the DPMM Tool

To gain qualitative insights into the needs, expectations, and experiences of HCPs with the tool, 19 HCPs—11 (58%) nurses and 8 (42%) physicians—were interviewed with open-ended questions at study end. Generally, expectations highlighted by HCPs were met or exceeded; improved efficiency and quality of patient care were the most prominent expectations of the tool and were mentioned by 8 out of 19 (42%) and 7 out of 19 (37%) interviewees, respectively (see Table 6 as well as Multimedia Appendix 6 for sample participant quotes). Improved efficiency and quality of patient care were also deemed the most value-adding attributes and were mentioned by 10 out of 19 (53%) interviewees for each attribute. Quotes highlighting improvements in efficiency and quality of patient care while using the tool are provided in Textboxes 1 and 2, respectively. Workload was the most prominent challenge of using the tool, as mentioned by 12 out of 19 (53%) interviewees, with some participants stating that the extra time needed to manage the standalone tool and enter the data sometimes compromised the efficiencies and time savings achieved elsewhere (see Table 6). Interoperability and system integration issues, as mentioned by 3 out of 19 (16%) interviewees but unlinked to tool functionality, were tightly related to workload challenges and could be considered the main cause for the perceived extra time (see Table 6).

Table 6. Expectations and perceived added value of the digital patient monitoring and management (DPMM) tool from qualitative health care professional interviews.

Expectations and perceptions of the DPMM tool	Expectation, perception, value, or challenge	Interviewees who mentioned the given theme (n=19)[a], n (%)
Expectations before the study		
Efficiency	Positive	8 (42)
Quality of care	Positive	7 (37)
Data generation	Positive	2 (11)
This is the future	Positive	2 (11)
Better patient education	Positive	1 (5)
More transparency about patients’ symptoms	Positive	1 (5)
Skepticism at the beginning	Neutral or negative	3 (16)
Expected more from the drug- and indication-specific cancer immunotherapy module	Neutral or negative	1 (5)
Perceived added value at the end of the study		
Efficiency	Key value attribute	10 (53)
Quality of care	Key value attribute	10 (53)
Communications and collaboration	Key value attribute	8 (42)
Workflow	Key value attribute	8 (42)
Empowerment	Key value attribute	5 (26)
Workload	Challenge	12 (63)
Interoperability and integration	Challenge	3 (16)

[a] Out of 19 interviewees, 11 (58%) were nurses and 8 (42%) were physicians.

Textbox 1. Quotes from two health care professionals (HCPs) to emphasize improvements in efficiency from using the digital patient monitoring and management tool.

- “When the patient came up with a problem, we were all prepared for it and the patient was there for a shorter amount of time because we already knew how to respond in advance. That was an improvement.” [HCP interviewee #13]
- “You could deal with problems beforehand. I had already seen what has been discussed, what she herself has said, and then you could just go from there. That then ultimately leads to shorter, more concise consultation times.” [HCP interviewee #19]

Textbox 2. Quotes from three health care professionals (HCPs) to emphasize improvements in quality of patient care from using the digital patient monitoring and management tool.

- “It became evident that we can use it for a more structured follow-up of the side effects of the treatments and symptoms of the patients.” [HCP interviewee #15]
- “For the data quality, for the care, it is a benefit. Undoubtedly. And they all felt really well cared-for, no doubt about it.” [HCP interviewee #13]
- “Perhaps some points are red [high severity] or newly yellow [medium severity], then you can say, ‘They must come to therapy earlier so that the doctor can talk to them directly and decide whether they would need only/either treatment or perhaps even hospitalization.’” [HCP interviewee #14]

Use of the DPMM Tool’s Individual Functions

Among HCPs who responded to the end-of-study survey, the most commonly appreciated functions of the tool were the patient symptom alerts (26/27, 96%) and the direct message communication function between patients and HCPs (19/27, 70%; see Table 7). Other important features for HCPs were the ability to use the tool during patient consultations (15/27, 56%), the facilitation of more effective conversations and referrals between nurses and physicians through the triage function

(13/27, 48%), the ability to use the tool during telephone consultations (13/27, 48%), and the onboarding of patients (13/27, 48%) (see Table 7).

Results from the end-of-study survey showed that patients had a similar preference with regard to the tool's functions; the most commonly appreciated functions were the symptoms questionnaire (20/21, 95%) and the direct message communication function between HCPs and patients (9/21, 43%). The patient card content—a PDF that could be completed digitally—offered to patients treated with atezolizumab was also appreciated; 2 of the 4 (50%) patients responding to the survey used the function (see Table 7). The median time to fill out the symptom questionnaire ranged from 2 minutes and 18 seconds (Clinic D, Germany) to 9 minutes and 56 seconds (Clinic J, Germany); the mean median time for questionnaire completion was 4 minutes and 14 seconds (see Multimedia Appendix 7).

Table 7. Functions of the digital patient monitoring and management (DPMM) tool most commonly appreciated by end-of-study survey respondents.

Functions of the DPMM tool	Nurses (n=19), n (%)	Physicians (n=8), n (%)	Patients (n=21), n (%)
Monitor the patient's symptoms	18 (95)	8 (100)	N/A[a]
Analyze the collected patient information	2 (11)	4 (50)	N/A
Draw patient reports	2 (11)	1 (13)	N/A
Directly communicate with the patients	14 (74)	5 (63)	N/A
Have more effective conversations and referrals between nurses and physicians	8 (42)	5 (63)	N/A
Use the tool during patient consultations	9 (47)	6 (75)	N/A
Use the tool during telephone consultations	7 (37)	6 (75)	N/A
Onboard patients to the tool	10 (53)	3 (38)	N/A
Symptoms questionnaire	N/A	N/A	20 (95)
Chat function	N/A	N/A	9 (43)
Patient card (n=4)[b]	N/A	N/A	2 (50)
Educational material[c]	N/A	N/A	3 (14)

[a] N/A: not applicable. This function was not relevant to this user group.

[b] Patient cards were part of the atezolizumab-specific material and were only relevant to patients treated with atezolizumab, of whom 4 responded to the survey.

[c] Educational material was offered to all 21 patients, although atezolizumab-specific educational material was offered only to atezolizumab-treated patients.

Impact of Individual Functions of the DPMM Tool on Users

Ratings from the end-of-study survey respondents demonstrated that the tool empowered patients, helping them to feel more in control (patient mean rating 3.9, SD 1.2), increasing their feelings of safety during their treatment (patient mean rating 3.9, SD 1.2), and helping them to feel more secure in evaluating their symptoms (patient mean rating 3.8, SD 1.3) (see Multimedia Appendix 8). HCPs appreciated the compact overview of patient development offered by the dashboard (HCP mean rating 4.25, SD 0.70) (see Multimedia Appendix 8).

Overall, according to in-module activities, the drug- and indication-specific educational material within the tool was engaged by, based on the number of downloads, 80% of patients (36/45) (see Multimedia Appendix 9), with two of the three atezolizumab-specific material items engaged by all 4 of the patients treated with atezolizumab who responded to the survey. Total median article viewing time across all clinics for the drug- and indication-specific educational material was approximately 3.5 hours; the longest viewing time was observed for the breathing exercises video, which was 1 hour, 28 minutes, and 56 seconds (see Multimedia Appendix 10). According to the end-of-study survey, the educational material was found by all users to be very helpful and informative, especially the lung cancer material (mean user rating 4.30, SD 0.73), the breathing exercises video (mean user rating 4.20, SD 0.93), and the CIT video (mean user rating 4.1, SD 0.7) (see Multimedia Appendix 8). The atezolizumab-specific material (ie, patient card, information on preparing for first infusion and treatment, and medication-specific material) received the highest rating of any of the materials offered; all 4 atezolizumab-treated patients who responded to the survey rated it as 5 (see Multimedia Appendix 8).

Data Sharing Statement

Qualified researchers may request access to analysis data via the corresponding author.

Discussion

Principal Findings

The CIT+ module that we co-developed with HCPs and patients was tested by physicians, nurses, and patients who considered themselves competent, proficient, or expert users. Most of them used the tool weekly (patients) or multiple times per week

(HCPs), and most used the tool for 10 minutes or less per session. As this was a pilot study, a population sample representative of the general cancer population was not its purpose.

Overall, the results of our proof-of-concept pilot study demonstrate that user acceptance of the tool was high, with usefulness and communication being the most appreciated attributes. Mean ratings were consistently over 3.5, which, similar to previous studies, were assumed to indicate high agreement [60]. A pilot study assessing the Diabetes Family Teamwork Online intervention in patients with type 1 diabetes also used a Likert scale, but a 7-point scale rather than a 5-point scale [61]. However, similarly, the study deemed approximately 70% of the maximum score to equal high feasibility [61]. In this study, the symptom questionnaire and symptom alerts were the most commonly used functions of the DPMM tool among patients and HCPs, respectively, followed by the direct message communication function and the drug- and indication-specific educational material. The symptom alert function was a key element, enabling HCPs to define alerts for particular symptoms and severity. HCPs stated that this enabled them to detect and manage critical symptoms earlier and personalize treatment plans, a result aligned with findings from other similar studies [24,62,63]. Our results highlight the essential features that a DPMM tool should include to better serve the needs of both patients and HCPs in clinical practice. Currently, few available tools combine all these features [20].

All survey respondents agreed that the attributes of the DPMM tool enabled more efficient and focused communication between patients and HCPs; this positive impact on communication has also been reported in previous studies [38,51,62]. Furthermore, the tool empowered patients, which has been shown to be correlated with an improved QoL [64,65], and helped them to evaluate and monitor their symptom progression.

In general, HCPs believed that the standalone tool was well integrated in their daily clinical workflow and improved efficiency within the health care team. This is based on positive insights from interviews (see Table 6), where HCPs reported improvements in quality of care and communications with the care team as well as time savings in patient visits with tool use, allowing for more clinically meaningful time with patients. Such positive insights are consistent with previous similar studies [66-68]. However, in this study, time savings were sometimes compromised by interoperability issues, such as lack of tool integration into the information technology (IT) system of the clinic and, consequently, data having to be gathered from both the web-based Kaiku Health platform and the clinic IT system. This challenge has been demonstrated in numerous other studies [38,40,69-71]. However, it should be resolved once the interoperability and system integration are incorporated beyond the pilot study as participating clinics and hospitals undertake complete rollout.

Notably, in addition to the time savings reported per patient visit by some HCPs, the tool also has the potential to free up time by decreasing the need for unscheduled outpatient appointments and telephone consultations, as reported by some patients (see Table 5). This is consistent with a possible reduction of scheduled visits in lung cancer patients following use of a DPMM tool that enables early detection of critical symptoms [23]. Considering most HCPs invested up to 30 minutes in introducing patients to the tool, and half of them saved 5 minutes or more per consultation, the time invested in patient onboarding was repaid within a few visits. The time saved through use of the DPMM tool can be invested in addressing other patient needs or serving more patients, highlighting the health-economic benefits of the tool.

Compared with studies of other ePRO-based tools, the CIT+ tool studied here showed similar high feasibility, patient engagement, and patient satisfaction [62,72-74]. However, the CIT+ module also harnesses features not commonly seen in other tools. For example, a 2019 systematic review of existing electronic symptom reporting systems developed for patients during cancer treatment found that fewer than half included a feature for delivering advice to patients in symptom self-management, and fewer than a third gave patients access to general educational information [20]. An even less common feature was the facility to support patient-HCP communication (15%) [20]. The CIT+ tool currently harnesses all these features, which is pertinent considering that previous research has indicated communication features, in particular, to be highly valued and utilized by patients [75-77].

Based on the results of this pilot study and insights from other studies, we propose several recommendations for future use in clinical practice or in study settings. For both, a positive and easy user experience is essential (eg, via optimization of the user interface), thereby enabling a choice of symptoms to be reported, as is providing automated contextual information according to user needs. In a clinical practice setting, a higher workflow efficiency and a much earlier detection of critical symptoms could be achieved with a seamless integration into the clinical workflow. Hence, integration into the local electronic health record system would be beneficial but may require substantial initial investments, given the complexities of existing systems. Such an integration may also enable local integration of patient self-management information and materials; these materials have been shown to be critical in supporting patient self-monitoring [77-79]. Integration may further enable connections to local resources and services, such as mental health or other supportive services, which can improve personalization of treatment according to patient needs and local standards. An enhanced local integration and tailoring to patient needs may improve adherence and allow patients to review and monitor their own data, which is the case for barely half of current ePRO systems [20].

In a clinical study setting, where assessment standards can be defined centrally and where stratified randomization of patients into standard-of-care control cohorts is possible, it is of interest to assess impact of the DPMM tool on patient outcomes and health care resource utilization and subsequently optimize for these elements. Regarding patient outcomes, assessment of QoL [22,23], in particular time to deterioration of symptoms and symptom development or adverse event intensity and duration over time, are important. Patients' abilities to work, their mental well-being, their self-care needs, and their self-efficacy are also important to assess. In regard to health care resource utilization, assessment of hospitalization days and emergency room visits [22], duration and adherence of treatment, co-medication use, and supportive care costs could yield valuable insights into the impact of a DPMM tool on clinical practice.

Limitations of the Study

A key limitation of this pilot study was its small sample size and single-arm design. Also, due to disease severity, the study was only able to recruit one participant for the patient advisory board. Further to this, DPMM tools are currently rarely used in routine clinical practice; as a result, there could be a bias for inclusion of sites, HCPs, and patients who are more accepting of these tools from the outset and are, therefore, more positive about their use. There are also several limitations associated with the chosen methods of data collection in this pilot study, namely surveys and interviews. Both rely on memory for answering questions, which may affect accuracy of the information received; furthermore, some bias may be created as participants who chose not to respond to survey questions or to participate in HCP interviews may have had different opinions from those who did so. Closed-ended questions, like those in our surveys, may have lower validity than open-ended questions, and answer options may be subject to the interpretation of different respondents. Finally, as discussed above, our DPMM tool was not integrated into the local electronic health record and patient management IT systems but was used as a standalone tool, which led to patient information needing to be gathered and recorded twice, causing additional workload for HCPs. This may have impeded HCPs from having an even more positive experience with the DPMM tool, thereby hampering their positive perceptions of the impact of the tool on clinical workflow efficiency. Hence, integration of the DPMM tool into the clinical IT data flow will be an important aspect of efficient routine clinical practice in the future. These limitations will be addressed in planned studies to evaluate the impact of the DPMM tool on patient health and health-economic benefits in a broader, better implemented, and more comprehensive approach.

Conclusions

Our results demonstrate high user satisfaction and acceptance of DPMM tools by HCPs and patients and highlight the contributions that DPMM tools can make to clinical care of patients with advanced or metastatic NSCLC treated with CIT monotherapy. Findings here will offer an incentive for continuous improvement and development of our tool, so that a platform can be provided that best serves the needs of HCPs and patients in this and other indications. The results add to the growing evidence base that DPMM tools can improve management of patients with cancer, empower patients, and have a health-economic impact by saving time in visits and reducing the need for patient telephone consultations [21,22]. Improvements in patient care have also been observed following the introduction of DPMM tools in other disease areas, such as multiple sclerosis [80-82]. Further studies or registries that allow investigation of the use of our DPMM tool may provide insights into whether its use would have any significant effect on other outcome measures, such as patient survival or QoL, or on health-economic benefits.

Acknowledgments

We would like to thank all the patients, HCPs, and investigators involved in this study. We would also like to thank Dr Sepiede Azghandi, Roche Pharma AG, Germany; Alexandra Schmid and Dr Flurina Pletscher, Roche Pharma AG, Switzerland; and Dr Anssi Linnankivi, Roche Oy, Finland, for their support with study setup and conduct. We would also like to thank Taina Hintikka, Roche Oy, Finland, and Claudia Hanselmann, Roche Pharma AG, Switzerland, for operational study support. This study was sponsored by F Hoffmann-La Roche Ltd, Basel, Switzerland. The study sponsor was involved in the study design, analysis and interpretation of data, writing of the article, and the decision to submit the manuscript for publication. Support for third-party writing assistance for this manuscript, furnished by Katie Wilson, PhD, and Stephen Salem, BSc, of Health Interactions, was provided by F Hoffmann-La Roche Ltd.

Authors' Contributions

OS contributed to the study concept and design; acquisition, analysis, or interpretation of data; and drafting of the manuscript. CJ is an external consultant for F Hoffmann-La Roche Ltd; CJ led the study design, was responsible for reporting, and contributed to the study concept; acquisition, analysis, or interpretation of data; drafting of the manuscript; and statistical analysis. JA contributed to acquisition, analysis, or interpretation of data and drafting of the manuscript. BL contributed to acquisition, analysis, or interpretation of data. SI contributed to the study concept and design; acquisition, analysis, or interpretation of data; and drafting of the manuscript. MK served as an external consultant for F Hoffmann-La Roche Ltd; MK was involved in the study operations, data collection, and reporting and contributed to the study concept and design; acquisition, analysis, or interpretation of data; drafting of the manuscript; and statistical analysis. JK contributed to acquisition, analysis, or interpretation of data and drafting of the manuscript. AK contributed to the study concept and design and the acquisition, analysis, or interpretation of data. RP contributed to acquisition, analysis, or interpretation of data. All authors contributed to critical revision of the manuscript for important intellectual content.

Conflicts of Interest

All authors received support for third-party writing assistance for this manuscript, provided by F Hoffmann-La Roche Ltd, Basel, Switzerland. OS and BL have acted in an advisory and consultancy role for F Hoffmann-La Roche Ltd (payment planned). CJ and MK have received honoraria from, and have acted as external consultants for, F Hoffmann-La Roche Ltd. JA and AK are employees of, and hold shares and stocks in, F Hoffmann-La Roche Ltd. SI has acted in a consultancy and advisory role for Bristol-Myers Squibb, Roche, and Merck Sharp & Dohme; has participated in a speaker bureau or provided expert testimony for Boehringer Ingelheim; is an employee at an institution that has received a research grant or funding from Roche; and has received travel and accommodation expenses from Boehringer Ingelheim, Merck Sharp & Dohme, Roche, Novartis, and Kaiku Health. MK is a part-time contractor for F Hoffmann-La Roche Ltd. JK has received honoraria from AstraZeneca, Bristol-Myers Squibb, Boehringer Ingelheim, Merck Sharp & Dohme, Novartis, Pfizer, Pierre Fabre, Roche, and Takeda; has acted in a consultancy and advisory role for AstraZeneca, Bristol-Myers Squibb, Boehringer Ingelheim, Faron, Kaiku Health, Merck Sharp & Dohme, Novartis, Pfizer, Pierre Fabre, Roche, and Takeda; has received a research grant or funding from Roche; and has received travel and accommodation expenses from AstraZeneca, Bristol-Myers Squibb, Boehringer Ingelheim, Faron, Kaiku Health, Merck Sharp & Dohme, Novartis, Pfizer, Pierre Fabre, Roche, and Takeda. MG is an employee at an institution that has received honoraria from F Hoffmann-La Roche Ltd. RP has acted in a consultancy and advisory role for Roche, Novartis, Merck Sharp & Dohme, Merck, Lilly, Bristol-Myers Squibb, AstraZeneca, Vifor Pharma, and Nutricia (all payments received by the institution) and is an employee at an institution that has received a research grant or funding from Roche, Novartis, Sanofi, and AbbVie.

Multimedia Appendix 1
Video clip showing how participants used the tool.
[MP4 File (MP4 Video), 31546 KB - jmir_v22i12e18655_app1.mp4]

Multimedia Appendix 2
The composition and size of the trained care teams.
[DOCX File , 14 KB - jmir_v22i12e18655_app2.docx]

Multimedia Appendix 3
Method of thematic analysis.
[DOCX File , 14 KB - jmir_v22i12e18655_app3.docx]

Multimedia Appendix 4
Preferred platform for tool usage in each of the user groups. The percentage indicates the absolute percentage distribution of all the log-ins for each device for each user group.
[PNG File , 42 KB - jmir_v22i12e18655_app4.png]

Multimedia Appendix 5
The number of chat messages sent between patients and health care professionals (HCPs) per clinic.
[DOCX File , 14 KB - jmir_v22i12e18655_app5.docx]

Multimedia Appendix 6
Participant quotes highlighting some of the main expectations before the pilot study.
[DOCX File , 14 KB - jmir_v22i12e18655_app6.docx]

Multimedia Appendix 7
Table showing median time to fill out the symptom questionnaire per clinic.
[DOCX File , 14 KB - jmir_v22i12e18655_app7.docx]

Multimedia Appendix 8
End-of-study survey respondents’ experiences of different features of the digital patient monitoring and management (DPMM) tool (N=48: 19 [40%] nurses, 8 [17%] physicians, and 21 [44%] patients). Data are the means of the given values on a scale of 1 (minimum value) to 5 (maximum value) for that user group.
[PNG File , 123 KB - jmir_v22i12e18655_app8.png]

Multimedia Appendix 9
Patient engagement in disease- and medication-specific educational material.
[DOCX File , 17 KB - jmir_v22i12e18655_app9.docx]

Multimedia Appendix 10
Median reading times of drug- and indication-specific educational material.

[DOCX File , 16 KB - jmir_v22i12e18655_app10.docx]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel R, Torre L, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Perez-Moreno P, Brambilla E, Thomas R, Soria J. Squamous cell carcinoma of the lung: Molecular subtypes and therapeutic opportunities. Clin Cancer Res 2012 May 01;18(9):2443-2451 [FREE Full text] [doi: 10.1158/1078-0432.CCR-11-2370] [Medline: 22407829]
3. National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®): Non-Small Cell Lung Cancer. Version 7.2019. 2019 Aug 30. URL: https://tinyurl.com/ydgkgjv5 [accessed 2020-12-06]
4. Planchard D, Popat S, Kerr K, Novello S, Smit EF, Faivre-Finn C, ESMO Guidelines Committee. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2018 Oct 01;29(Suppl 4):iv192-iv237 [FREE Full text] [doi: 10.1093/annonc/mdy275] [Medline: 30285222]
5. Roche Registration GmbH. Annex 1: Summary of product characteristics. European Medicines Agency. 2020. URL: https://www.ema.europa.eu/en/documents/product-information/tecentriq-epar-product-information_en.pdf [accessed 2020-02-01]
6. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, KEYNOTE-189 Investigators. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med 2018 May 31;378(22):2078-2092. [doi: 10.1056/NEJMoa1801005] [Medline: 29658856]
7. Paz-Ares L, Luft A, Vicente D, Tafreshi A, Gümüş M, Mazières J, KEYNOTE-407 Investigators. Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer. N Engl J Med 2018 Nov 22;379(21):2040-2051. [doi: 10.1056/NEJMoa1810865] [Medline: 30280635]
8. Fehrenbacher L, Spira A, Ballinger M, Kowanetz M, Vansteenkiste J, Mazieres J, POPLAR Study Group. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): A multicentre, open-label, phase 2 randomised controlled trial. Lancet 2016 Apr 30;387(10030):1837-1846. [doi: 10.1016/S0140-6736(16)00587-0] [Medline: 26970723]
9. Herbst RS, Baas P, Kim D, Felip E, Pérez-Gracia JL, Han J, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): A randomised controlled trial. Lancet 2016 Apr 09;387(10027):1540-1550. [doi: 10.1016/S0140-6736(15)01281-7] [Medline: 26712084]
10. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, OAK Study Group. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): A phase 3, open-label, multicentre randomised controlled trial. Lancet 2017 Jan 21;389(10066):255-265 [FREE Full text] [doi: 10.1016/S0140-6736(16)32517-X] [Medline: 27979383]
11. Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, IMpower150 Study Group. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. N Engl J Med 2018 Jun 14;378(24):2288-2301. [doi: 10.1056/NEJMoa1716948] [Medline: 29863955]
12. West H, McCleod M, Hussein M, Morabito A, Rittmeyer A, Conter HJ, et al. Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): A multicentre, randomised, open-label, phase 3 trial. Lancet Oncol 2019 Jul;20(7):924-937. [doi: 10.1016/S1470-2045(19)30167-6] [Medline: 31122901]
13. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N Engl J Med 2015 Oct 22;373(17):1627-1639 [FREE Full text] [doi: 10.1056/NEJMoa1507643] [Medline: 26412456]
14. Immunotherapy Side Effects. Lugano, Switzerland: European Society for Medical Oncology (ESMO); 2017. URL: https://www.esmo.org/content/download/124130/2352601/file/ESMO-Patient-Guide-on-Immunotherapy-Side-Effects.pdf [accessed 2020-02-01]
15. Iyer S, Roughley A, Rider A, Taylor-Stokes G. The symptom burden of non-small cell lung cancer in the USA: A real-world cross-sectional study. Support Care Cancer 2014 Jan;22(1):181-187. [doi: 10.1007/s00520-013-1959-4] [Medline: 24026981]
16. Andikyan V, Rezk Y, Einstein MH, Gualtiere G, Leitao MM, Sonoda Y, et al. A prospective study of the feasibility and acceptability of a web-based, electronic patient-reported outcome system in assessing patient recovery after major gynecologic cancer surgery. Gynecol Oncol 2012 Nov;127(2):273-277 [FREE Full text] [doi: 10.1016/j.ygyno.2012.07.124] [Medline: 22871467]
17. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 2013;13:211 [FREE Full text] [doi: 10.1186/1472-6963-13-211] [Medline: 23758898]
18. Velikova G, Keding A, Harley C, Cocks K, Booth L, Smith AB, et al. Patients report improvements in continuity of care when quality of life assessments are used routinely in oncology practice: Secondary outcomes of a randomised controlled trial. Eur J Cancer 2010 Sep;46(13):2381-2388. [doi: 10.1016/j.ejca.2010.04.030] [Medline: 20570138]

19. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst 2014 Sep;106(9):dju244 [FREE Full text] [doi: 10.1093/jnci/dju244] [Medline: 25265940]
20. Warrington L, Absolom K, Conner M, Kellar I, Clayton B, Ayres M, et al. Electronic systems for patients to report and manage side effects of cancer treatment: Systematic review. J Med Internet Res 2019 Jan 24;21(1):e10875 [FREE Full text] [doi: 10.2196/10875] [Medline: 30679145]
21. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
22. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Jul 11;318(2):197-198. [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
23. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst 2017 Sep 01;109(9):1. [doi: 10.1093/jnci/djx029] [Medline: 28423407]
24. Denis F, Basch E, Septans A, Bennouna J, Urban T, Dueck AC, et al. Two-year survival comparing web-based symptom monitoring vs routine surveillance following treatment for lung cancer. JAMA 2019 Jan 22;321(3):306-307 [FREE Full text] [doi: 10.1001/jama.2018.18085] [Medline: 30667494]
25. Bakitas MA, Tosteson TD, Li Z, Lyons KD, Hull JG, Li Z, et al. Early versus delayed initiation of concurrent palliative oncology care: Patient outcomes in the ENABLE III randomized controlled trial. J Clin Oncol 2015 May 01;33(13):1438-1445 [FREE Full text] [doi: 10.1200/JCO.2014.58.6362] [Medline: 25800768]
26. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. London, UK: SAGE Publishing; Mar 2013.
27. Cresswell J, Poth C. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Fourth edition. Thousand Oaks, CA: SAGE Publishing; 2018.
28. Haanen J, Carbonnel F, Robert C, Kerr K, Peters S, Larkin J, ESMO Guidelines Committee. Management of toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2017 Jul 01;28(suppl_4):iv119-iv142 [FREE Full text] [doi: 10.1093/annonc/mdx225] [Medline: 28881921]
29. National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for Supportive Care: Cancer-Related Fatigue. 2020. URL: https://www.nccn.org/professionals/physician_gls/ [accessed 2020-12-06]
30. Absolom K, Holch P, Warrington L, Samy F, Hulme C, Hewison J, eRAPID Systemic Treatment Work Group. Electronic patient self-Reporting of Adverse-events: Patient Information and aDvice (eRAPID): A randomised controlled trial in systemic cancer treatment. BMC Cancer 2017 May 08;17(1):318 [FREE Full text] [doi: 10.1186/s12885-017-3303-8] [Medline: 28482877]
31. Gagnon MP, Orruño E, Asua J, Abdeljelil AB, Emparanza J. Using a modified technology acceptance model to evaluate healthcare professionals' adoption of a new telemonitoring system. Telemed J E Health 2012;18(1):54-59 [FREE Full text] [doi: 10.1089/tmj.2011.0066] [Medline: 22082108]
32. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
33. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
34. Holden RJ, Karsh B. The technology acceptance model: Its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
35. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: Systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
36. Garavand A, Mohseni M, Asadi H, Etemadi M, Moradi-Joo M, Moosavi A. Factors influencing the adoption of health information technologies: A systematic review. Electron Physician 2016 Aug;8(8):2713-2718 [FREE Full text] [doi: 10.19082/2713] [Medline: 27757179]
37. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: A qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
38. Bidmead E, Marshall A. A case study of stakeholder perceptions of patient held records: The Patients Know Best (PKB) solution. Digit Health 2016;2:e18072 [FREE Full text] [doi: 10.1177/2055207616668431] [Medline: 29942567]
39. Bishop TF, Press MJ, Mendelsohn JL, Casalino LP. Electronic communication improves access, but barriers to its widespread adoption remain. Health Aff (Millwood) 2013 Aug;32(8):1361-1367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1151] [Medline: 23918479]

40. Chung C, Cook J, Bales E, Zia J, Munson SA. More than telemonitoring: Health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. J Med Internet Res 2015 Aug 21;17(8):e203 [FREE Full text] [doi: 10.2196/jmir.4364] [Medline: 26297627]
41. Putzer GJ, Park Y. Are physicians likely to adopt emerging mobile technologies? Attitudes and innovation factors affecting smartphone use in the Southeastern United States. Perspect Health Inf Manag 2012;9:1b [FREE Full text] [Medline: 22737094]
42. Anderson K, Francis T, Ibanez-Carrasco F, Globerman J. Physician's perceptions of telemedicine in HIV care provision: A cross-sectional web-based survey. JMIR Public Health Surveill 2017 May 30;3(2):e31 [FREE Full text] [doi: 10.2196/publichealth.6896] [Medline: 28559226]
43. de Souza CHA, Morbeck RA, Steinman M, Hors CP, Bracco MM, Kozasa EH, et al. Barriers and benefits in telemedicine arising between a high-technology hospital service provider and remote public healthcare units: A qualitative study in Brazil. Telemed J E Health 2017 Jun;23(6):527-532. [doi: 10.1089/tmj.2016.0158] [Medline: 27911667]
44. Esterle L, Mathieu-Fritz A. Teleconsultation in geriatrics: Impact on professional practice. Int J Med Inform 2013 Aug;82(8):684-695. [doi: 10.1016/j.ijmedinf.2013.04.006] [Medline: 23746716]
45. Hickson R, Talbert J, Thornbury WC, Perin NR, Goodin AJ. Online medical care: The current state of "eVisits" in acute primary care delivery. Telemed J E Health 2015 Feb;21(2):90-96. [doi: 10.1089/tmj.2014.0022] [Medline: 25474083]
46. Vest BM, Hall VM, Kahn LS, Heider AR, Maloney N, Singh R. Nurse perspectives on the implementation of routine telemonitoring for high-risk diabetes patients in a primary care setting. Prim Health Care Res Dev 2017 Jan;18(1):3-13. [doi: 10.1017/S1463423616000190] [Medline: 27269513]
47. Catan G, Espanha R, Mendes RV, Toren O, Chinitz D. Health information technology implementation - Impacts and policy considerations: A comparison between Israel and Portugal. Isr J Health Policy Res 2015;4:41 [FREE Full text] [doi: 10.1186/s13584-015-0040-9] [Medline: 26269740]
48. Cox A, Illsley M, Knibb W, Lucas C, O'Driscoll M, Potter C, et al. The acceptability of e-technology to monitor and assess patient symptoms following palliative radiotherapy for lung cancer. Palliat Med 2011 Oct;25(7):675-681. [doi: 10.1177/0269216311399489] [Medline: 21474620]
49. L'Esperance ST, Perry DJ. Assessing advantages and barriers to telemedicine adoption in the practice setting: A MyCareTeam(TM) exemplar. J Am Assoc Nurse Pract 2016 Jun;28(6):311-319. [doi: 10.1002/2327-6924.12280] [Medline: 26178417]
50. MacNeill V, Sanders C, Fitzpatrick R, Hendy J, Barlow J, Knapp M, et al. Experiences of front-line health professionals in the delivery of telehealth: A qualitative study. Br J Gen Pract 2014 Jul;64(624):e401-e407 [FREE Full text] [doi: 10.3399/bjgp14X680485] [Medline: 24982492]
51. Mileski M, Kruse CS, Catalani J, Haderer T. Adopting telemedicine for the self-management of hypertension: Systematic review. JMIR Med Inform 2017 Oct 24;5(4):e41 [FREE Full text] [doi: 10.2196/medinform.6603] [Medline: 29066424]
52. Odeh B, Kayyali R, Nabhani-Gebara S, Philip N. Implementing a telehealth service: Nurses' perceptions and experiences. Br J Nurs 2014;23(21):1133-1137. [doi: 10.12968/bjon.2014.23.21.1133] [Medline: 25426527]
53. Schneider T, Panzera AD, Martinasek M, McDermott R, Couluris M, Lindenberger J, et al. Physicians' perceptions of mobile technology for enhancing asthma care for youth. J Child Health Care 2016 Jun;20(2):153-163. [doi: 10.1177/1367493514556555] [Medline: 25427556]
54. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: A systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
55. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and facilitators to eHealth use in daily practice: Perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 05;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
56. Duhm J, Fleischmann R, Schmidt S, Hupperts H, Brandt SA. Mobile electronic medical records promote workflow: Physicians' perspective from a survey. JMIR Mhealth Uhealth 2016 Jun 06;4(2):e70 [FREE Full text] [doi: 10.2196/mhealth.5464] [Medline: 27268720]
57. Ehrler F, Ducloux P, Wu DTY, Lovis C, Blondon K. Acceptance of a mobile application supporting nurses workflow at patient bedside: Results from a pilot study. Stud Health Technol Inform 2018;247:506-510. [Medline: 29678012]
58. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
59. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
60. Riad Mousa Jaradat MI, Moh'd Ali Smadi Z. Applying the technology acceptance model to the introduction of mobile healthcare information systems. Int J Behav Healthc Res 2013 Jan;4(2):123-143. [doi: 10.1504/ijbhr.2013.057363]
61. Thompson D, Callender C, Gonynor C, Cullen KW, Redondo MJ, Butler A, et al. Using relational agents to promote family communication around type 1 diabetes self-management in the diabetes family teamwork online intervention: Longitudinal pilot study. J Med Internet Res 2019 Sep 13;21(9):e15318 [FREE Full text] [doi: 10.2196/15318] [Medline: 31538940]
62. Maguire R, Ream E, Richardson A, Connaghan J, Johnston B, Kotronoulas G, et al. Development of a novel remote patient monitoring system: The advanced symptom management system for radiotherapy to improve the symptom experience of

patients with lung cancer receiving radiotherapy. Cancer Nurs 2015;38(2):E37-E47. [doi: 10.1097/NCC.0000000000000150] [Medline: 24836956]

63. Howell D, Molloy S, Wilkinson K, Green E, Orchard K, Wang K, et al. Patient-reported outcomes in routine cancer clinical practice: A scoping review of use, impact on health outcomes, and implementation factors. Ann Oncol 2015 Sep;26(9):1846-1858 [FREE Full text] [doi: 10.1093/annonc/mdv181] [Medline: 25888610]
64. Hibbard JH, Mahoney E, Sonet E. Does patient activation level affect the cancer patient journey? Patient Educ Couns 2017 Jul;100(7):1276-1279. [doi: 10.1016/j.pec.2017.03.019] [Medline: 28330715]
65. Papadopoulou C, Kotronoulas G, Schneider A, Miller MI, McBride J, Polly Z, et al. Patient-reported self-efficacy, anxiety, and health-related quality of life during chemotherapy: Results from a longitudinal study. Oncol Nurs Forum 2017 Jan 01;44(1):127-136. [doi: 10.1188/17.ONF.127-136] [Medline: 27991612]
66. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians' role in the adoption of an oncology decision support app in Europe and its implications for organizational practices: Qualitative case study. JMIR Mhealth Uhealth 2019 May 03;7(5):e13555 [FREE Full text] [doi: 10.2196/13555] [Medline: 31066710]
67. Mueller KJ, Potter AJ, MacKinney AC, Ward MM. Lessons from tele-emergency: Improving care quality and health outcomes by expanding support for rural care systems. Health Aff (Millwood) 2014 Feb;33(2):228-234. [doi: 10.1377/hlthaff.2013.1016] [Medline: 24493765]
68. Ruiz Morilla MD, Sans M, Casasa A, Giménez N. Implementing technology in healthcare: Insights from physicians. BMC Med Inform Decis Mak 2017 Jun 27;17(1):92 [FREE Full text] [doi: 10.1186/s12911-017-0489-2] [Medline: 28655299]
69. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: Systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
70. Morrow S, Daines L, Wiener-Ogilvie S, Steed L, McKee L, Caress A, et al. Exploring the perspectives of clinical professionals and support staff on implementing supported self-management for asthma in UK general practice: An IMPART qualitative study. NPJ Prim Care Respir Med 2017 Jul 18;27(1):45 [FREE Full text] [doi: 10.1038/s41533-017-0041-y] [Medline: 28720751]
71. Öberg U, Orre CJ, Isaksson U, Schimmer R, Larsson H, Hörnsten Å. Swedish primary healthcare nurses' perceptions of using digital eHealth services in support of patient self-management. Scand J Caring Sci 2018 Jun;32(2):961-970. [doi: 10.1111/scs.12534] [Medline: 28960451]
72. Chan M, Ang E, Duong MC, Chow YL. An online symptom care and management system to monitor and support patients receiving chemotherapy: A pilot study. Int J Nurs Pract 2013 Feb;19 Suppl 1:14-18. [doi: 10.1111/ijn.12020] [Medline: 23425375]
73. Peltola MK, Lehikoinen JS, Sippola LT, Saarilahti K, Mäkitie AA. A novel digital patient-reported outcome platform for head and neck oncology patients-A pilot study. Clin Med Insights Ear Nose Throat 2016;9:1-6 [FREE Full text] [doi: 10.4137/CMENT.S40219] [Medline: 27721662]
74. Tran C, Dicker A, Leiby B, Gressen E, Williams N, Jim H. Utilizing digital health to collect electronic patient-reported outcomes in prostate cancer: Single-arm pilot trial. J Med Internet Res 2020 Mar 25;22(3):e12689 [FREE Full text] [doi: 10.2196/12689] [Medline: 32209536]
75. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of internet-based asynchronous communication between health providers and patients with a chronic condition: A systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
76. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 06;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
77. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: A systematic review. J Med Internet Res 2013 Feb 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
78. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949. [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]
79. McCorkle R, Ercolano E, Lazenby M, Schulman-Green D, Schilling LS, Lorig K, et al. Self-management: Enabling and empowering patients living with cancer as a chronic illness. CA Cancer J Clin 2011;61(1):50-62 [FREE Full text] [doi: 10.3322/caac.20093] [Medline: 21205833]
80. Montalban X, Mulero P, Midaglia L, Graves J, Hauser S, Julian L, et al. FLOODLIGHT: Remote self-monitoring is accepted by patients and provides meaningful, continuous sensor-based outcomes consistent with and augmenting conventional in-clinic measures. Neurology 2018 Apr;90(15 Supplement):P4.382.
81. Montalban X, Mulero P, Midaglia L, Graves J, Hauser S, Julian L, et al. FLOODLIGHT: Smartphone-based self-monitoring is accepted by patients and provides meaningful, continuous digital outcomes augmenting conventional in-clinic multiple sclerosis measures. Neurology 2019;92(15 Supplement):P3.2-024.
82. Midaglia L, Mulero P, Montalban X, Graves J, Hauser SL, Julian L, et al. Adherence and satisfaction of smartphone- and smartwatch-based remote active testing and passive monitoring in people with multiple sclerosis: Nonrandomized

interventional feasibility study. J Med Internet Res 2019 Aug 30;21(8):e14863 [FREE Full text] [doi: 10.2196/14863] [Medline: 31471961]

Abbreviations

ALK: anaplastic lymphoma kinase
BRAF: B-Raf proto-oncogene
CIT: cancer immunotherapy
CIT+: drug- and indication-specific cancer immunotherapy
CTCAE: Common Terminology Criteria for Adverse Events
DPMM: digital patient monitoring and management
EGFR: epidermal growth factor receptor
ePRO: electronic patient-reported outcome
HCP: health care professional
IT: information technology
MET: N-methyl-N'-nitroso-guanidine human osteosarcoma transforming
NCI: National Cancer Institute
NSCLC: non-small cell lung cancer
NTRK: neurotrophic tropomyosin receptor kinase
OS: overall survival
PD-L1: programmed death-ligand 1
PRO: patient-reported outcome
QoL: quality of life
RET: rearranged during transfection
ROS1: ROS proto-oncogene 1
TAM: technology acceptance model

Edited by G Eysenbach, R Kukafka; submitted 19.03.20; peer-reviewed by Y Cheng, G Kotronoulas, S Nabhani-Gebara, B Arents; comments to author 06.04.20; revised version received 01.06.20; accepted 02.11.20; published 21.12.20.

Please cite as:
Schmalz O, Jacob C, Ammann J, Liss B, Iivanainen S, Kammermann M, Koivunen J, Klein A, Popescu RA
Digital Monitoring and Management of Patients With Advanced or Metastatic Non-Small Cell Lung Cancer Treated With Cancer Immunotherapy and Its Impact on Quality of Clinical Care: Interview and Survey Study Among Health Care Professionals and Patients
J Med Internet Res 2020;22(12):e18655
URL: http://www.jmir.org/2020/12/e18655/
doi:10.2196/18655
PMID:33346738

Original Paper

Evaluation of Four Artificial Intelligence–Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study

Aleksandar Ćirković, MD, MSc

Corresponding Author:
Aleksandar Ćirković, MD, MSc
Schulgasse 21
Weiden, 92637
Germany
Phone: 49 1788603753
Email: aleksandar.cirkovic@mailbox.org

Abstract

Background: Consumer-oriented mobile self-diagnosis apps have been developed using undisclosed algorithms, presumably based on machine learning and other artificial intelligence (AI) technologies. The US Food and Drug Administration now discerns apps with learning AI algorithms from those with stable ones and treats the former as medical devices. To the author's knowledge, no self-diagnosis app testing has been performed in the field of ophthalmology so far.

Objective: The objective of this study was to test apps that were previously mentioned in the scientific literature on a set of diagnoses in a deliberate time interval, comparing the results and looking for differences that hint at "nonlocked" learning algorithms.

Methods: Four apps from the literature were chosen (Ada, Babylon, Buoy, and Your.MD). A set of three ophthalmology diagnoses (glaucoma, retinal tear, dry eye syndrome) representing three levels of urgency was used to simultaneously test the apps' diagnostic efficiency and treatment recommendations in this specialty. Two years was the chosen time interval between the tests (2018 and 2020). Scores were awarded by one evaluating physician using a defined scheme.

Results: Two apps (Ada and Your.MD) received significantly higher scores than the other two. All apps either worsened in their results between 2018 and 2020 or remained unchanged at a low level. The variation in the results over time indicates "nonlocked" learning algorithms using AI technologies. None of the apps provided correct diagnoses and treatment recommendations for all three diagnoses in 2020. Two apps (Babylon and Your.MD) asked significantly fewer questions than the other two ($P<.001$).

Conclusions: "Nonlocked" algorithms are used by self-diagnosis apps. The diagnostic efficiency of the tested apps seems to worsen over time, with some apps being more capable than others. Systematic studies on a wider scale are necessary for health care providers and patients to correctly assess the safety and efficacy of such apps and for correct classification by health care regulating authorities.

(J Med Internet Res 2020;22(12):e18097) doi:10.2196/18097

KEYWORDS

artificial intelligence; machine learning; mobile apps; medical diagnosis; mHealth

Introduction

Algorithms and machine learning (ML) have conquered, transformed, and essentially revolutionized people's everyday lives in many aspects (eg, with personalized Google searches, self-driving cars, or convenient smartphone apps) [1,2]. In recent years, self-diagnosis apps have emerged that allow patients to look for a diagnosis based on entered symptoms [3,4]. ML is already being standardly used in various applications like estimating diagnoses from radiology images [5], but the adoption and acceptance of new technologies in health care in general is curbed by trust issues, strict regulations, and lack of thorough investigation [3,6]. Little testing of the aforementioned apps has been previously performed; Semigran et al tested self-diagnosis apps in general in 2015 but did not mention ML as an underlying technology [7]. A recent scoping review by Aboueid et al in 2019 named several apps of this type [3], but only two have been tested in their diagnostic functionality so far [8-10]. The US Food and Drug Administration (FDA) had previously excluded "symptom checker" apps from the enforcement of the strict rules that are usually applied to medical devices [11] but has lately released a white paper with a proposal for possible changes in the regulation of self-diagnosis apps, introducing a new discrimination between "locked" and artificial intelligence (AI)–based/ML-based learning algorithms, with the latter falling under a stricter set of rules [12]. However, the provided definitions of the two categories are still vague and rely on the manufacturer admitting the use of AI-based technologies rather than on hard criteria. Repeated examinations of such apps would offer insight into whether the results they offer are stable over time or change, which would also indicate actual use of "learning" AI technologies by the manufacturers, rendering the apps more than plain symptom checkers. Thus, the first aim of this study was to test apps that have previously been mentioned in the scientific literature on a set of diagnoses and to subsequently follow up on the results after a deliberate time interval to determine whether the algorithms change over time. A significant change in the results might indicate that learning algorithms are used by the manufacturers.

None of the aforementioned studies had tested self-diagnosis apps in the field of ophthalmology, the author's primary specialty with over 12 years of clinical experience. Thus, the second aim of the study was to test the efficacy of self-diagnosis apps by presenting them with three common diagnoses representing three levels of urgency in this field to see if their results and treatment recommendations match those of someone familiar with the topic.

Methods

Overview

The mobile apps were tested on Android 9 and Android 10, and the web app was tested on Google Chrome for OSX. All tests took place in Germany, and in all cases the English user interfaces were used. For all apps, up-to-date versions in the Google Play Store or on the internet were used; they correspond to the dates noted for each of the diagnostic walk-throughs (see Multimedia Appendix 1). The programs that were included in the testing were those that have been mentioned in the scientific literature: Ada [3,8,13], Babylon Health or Babylon Check [3,9,13], Buoy Health [14], and Your.MD [13]. Excluded from testing were Baidu Doctor [15], which is available only in Chinese, and K Health [3], which could not be downloaded onto the devices due to regional unavailability on iOS and unresolvable compatibility issues on Android, which may also stem from regional restriction without explicit notification. While there are reports about Buoy that claim that an Android-based app is in development, at the time of testing it could only be used as a web app [16]. The basic functional principles as understood by the author or described by the software developers themselves and recent literature, including gray literature where scientific literature was unavailable, were summarized. There is no consensus on broad testing of AI-based apps yet, but in the past, symptom checker apps have been tested using a randomized set of virtual diagnoses combined with gathered patient information in sets called vignettes [7]. The same setup has been used recently in a 2018 study by Razzaki et al to test Babylon Health [10] and in a 2019 study by Jungmann et al to test Ada Health in the field of mental disorders [8]. Only one physician creating one virtual patient per diagnosis was involved in this work; thus, an abbreviated and simplified version of this procedure was performed as follows. Three defined diagnoses from the ophthalmology branch of medicine were entered via the apps' given user interfaces: representing an absolute emergency (immediate treatment recommended), a glaucoma attack in one eye with the typical combination of a painful red eye for about two hours, blurred vision, a headache, and other symptoms depending on how each app asked its questions (see Multimedia Appendix 1 for the walk-through of all diagnoses as experienced); as a relative emergency (same-day treatment recommended), a retinal tear; and as a diagnosis that does not need immediate treatment and may also be self-treated first, dry eyes. For all three diagnoses, there are no strict general clinical therapy guidelines available, but some regional societies offer general recommendations [17-19]. The symptoms of all three diagnoses are common knowledge in ophthalmology and are thoroughly examined in the American Academy of Ophthalmology's Basic and Clinical Science Course, congruent with the author's applied knowledge on the cases [20]. As no clear guidelines exist, a minimum requirement for the apps was to not underestimate the urgency of the patient's situation, regardless of whether a diagnosis was found or not. The foremost diagnoses and treatment recommendations given by the apps were assessed by the author and a score was awarded: for a correct diagnosis/treatment, 1 point; for a partially correct diagnosis or treatment, half a point, which was awarded if a correct diagnosis or treatment was not provided, but the answer would not mislead the user or underestimate the urgency (eg, if no diagnosis was found, and the app recommended visiting a real physician); for all diagnoses/treatments that did not meet these requirements, 0 points. More specifically, for glaucoma, anything less than emergency treatment would result in 0 points; for the retinal tear diagnosis, 1 point was awarded to treatment recommendation ranges of "instantly" to very few days; and for dry eyes, 0 points were awarded if treatment was deemed urgent, and half a point was awarded if trying self-treatment was not recommended before seeing a physician. A virtual anonymous patient was created to be diagnosed to prevent the potential influence of phone-based data (phone type, GPS coordinates, country, etc). Roughly two years was chosen as a deliberate time interval between tests, based on the assumption of a slow but continuous rise of the apps' user count, and thus a slow but continuous buildup of internal data to process for possible improvements of presumed learning algorithms used by the apps. Other than the author, there were no human subjects involved in the process of this research. *P* values were calculated with a Student *t* test for independent variables using SPSS (version 16.0; IBM Corp).

Ada

Ada is a Berlin-based app that was first tested and gained popularity on the New Zealand market in 2016 and was released more broadly afterward [21,22]. It uses a chat bot to collect data from the user, selecting symptoms from a list generated in response to the user's free text input and subsequently asking questions that adapt to previously entered information. The resulting report may then be sent to a physician on behalf of the user.

Babylon

Babylon is a London-based app that primarily focuses on the UK market. It started out in 2013 as a service provider for online consultations with real-life physicians, and since 2016 it has added a chat bot that presents the user with simple or multiple-choice questions for symptom assessment [23]. An explicit description of the ML algorithms that are used is not available, but judging by the publicly available information from gray literature, the use of recurrent neural networks (RNNs) for deep learning may be involved, and Python may be used as the primary programming language [24]. Ni et al mentioned the use of Bayesian networks, although a source was not given [25].

Buoy Health

Buoy has been developed at Harvard Medical School since 2014 as a smart symptom checker with an undisclosed algorithm, supposedly relying on natural language processing (NLP)–extracted data from 18,000 clinical papers [26]. As stated by its chief executive officer and founder, Buoy Health specifically does not use decision trees but "dynamically picks" 1 of 30,000 questions based on the principle of greatest reduction of diagnostic uncertainty, which does not necessarily imply the use of neural networks. By his accounts, its diagnostic certainty is within a range of 90.9%-98%, without a detailed explanation [27].

Your.MD

Your.MD was founded in 2012 in Oslo, Norway, and is now based in London [28,29]. Users ask free questions via its chat bot and it in turn presents simple or multiple-choice questions. The algorithms that are used are undisclosed, but judging by the publicly available information, Python may be used as a primary programming language and, according to its CEO, Bayesian networks may also be used [30].

Results

Ada

Ada diagnosed the angle closure glaucoma correctly in 2018 but misdiagnosed it as cluster headache in 2020 without mentioning glaucoma (this result and all following are shown in Table 1 and Table 2; for the walk-throughs with raw output data for all apps, see Multimedia Appendix 1; for an overview of the relevant results and awarded scores, see Multimedia Appendix 2).

Table 1. Testing of all four apps in three virtual patients with different diagnoses in 2018. Points awarded for diagnosis/treatment: (-)=0, (*)=0.5, (•)=1.

Diagnosis	Glaucoma	Retinal tear	Dry eyes
Ada	(•)/(•)	(•)/(•)	(•)/(-)
Babylon	(-)/(•)	(-)/(-)	(-)/(-)
Buoy	(-)/(•)	(-)/(-)	(-)/(*)
Your.MD	(•)/(•)	(•)/(•)	(•)/(•)

Table 2. Testing of all four apps in three virtual patients with different diagnoses in 2020. Points awarded for diagnosis/treatment: (-)=0, (*)=0.5, (•)=1.

Diagnosis	Glaucoma	Retinal tear	Dry eyes
Ada	(-)/(-)	(•)/(•)	(•)/(*)
Babylon	(-)/(•)	(-)/(-)	(-)/(-)
Buoy	(-)/(-)	(-)/(•)	(-)/(*)
Your.MD	(•)/(•)	(-)/(•)	(*)/(-)

Babylon

In the glaucoma attack case, in both years Babylon recommended seeking emergency treatment after five questions when the user classified the pain as "severe," with no further statement about a diagnosis. The retinal tear was not diagnosed due to "insufficient information," and Babylon recommended referral to an online physician or real-life general practitioner. In the dry eyes case, Babylon did not state a diagnosis either and classified this as a relative emergency (same-day medical treatment recommended). There was no change in 2020.

Buoy Health

Buoy yielded a correct diagnosis neither in 2018—although its second suggestion of "Blepharitis" could be interpreted as partly correct in the dry eyes case [31]—nor in 2020; the result for the retinal tear inquiry was far off in 2018 with "Cataract" or "Bone disease" given as possible causes. The efficiency and accuracy did not improve in 2020.

Your.MD

Your.MD was able to output the correct diagnosis in all three tests in 2018, requiring distinctly fewer questions. In contrast to Ada, the treatment priorities were categorized correctly in all three, recognizing dry eyes as self-treatable. It was the only app to correctly identify the angle closure glaucoma, which it did in 2020; in 2018, it had only stated "Glaucoma". In case of the retinal tear, however, it was unable to correctly identify it in 2020, while it had done so in 2018. In the dry eyes case, it changed the advice from self-treatment in 2018, which is correct, to an emergency in 2020, which is not.

Technologies

All tested apps require an online connection to use the diagnosis function. All rely on a chat bot that is likely based on NLP and subsequently on discrete answers to questions to process user input, but they significantly differ on how the information is treated, which questions are asked, and which conclusions are drawn from the information (see Multimedia Appendix 1). There is no substantial information available on the algorithms used by the apps.

Summary of Comparison Between Performances in 2018 and 2020

The average number of questions changed from 27.3 for Ada, 11 for Babylon, 31.3 for Buoy, and 10 for Your.MD in 2018 to 31 for Ada (*P*=.38), 9 for Babylon (*P*=.64), 30.3 for Buoy (*P*=.63), and 10.3 for Your.MD (*P*=.84) in 2020 (see Multimedia Appendix 3). In the average number of questions asked in both time periods, no difference could be found between Ada and Buoy (*P*=.41) and between Babylon and Your.MD (*P*=.93), but significant differences were found between Ada and Babylon (*P*<.001), Ada and Your.MD (*P*<.001), Babylon and Buoy (*P*<.001), and Buoy and Your.MD (*P*<.001). The scores from 2018 to 2020 changed in Ada from 3/2 to 2/1.5 (*P*=.37/.73) and in Your.MD from 3/3 to 1.5/2 (*P*=.16/.37), while Babylon and Buoy remained unchanged at 0/1 and 0/1.5, respectively (Table 1 and Table 2).

The average scores were 2.5/1.75 for Ada, 0/1 for Babylon, 0/1.5 for Buoy, and 2.25/2.5 for Your.MD, and the sums over both years were 5/3.5 for Ada, 0/2 for Babylon, 0/3 for Buoy, and 4.5/5 for Your.MD. In the sum of all points, Ada and Your.MD (*P*=.70) and Babylon and Buoy (*P*=.56) did not differ significantly, while Ada and Babylon (*P*=.02), Ada and Buoy (*P*=.03), Babylon and Your.MD (*P*=.01) and Buoy and Your.MD (*P*=.01) did.

Discussion

During testing of the apps, some notable observations could be made about their behaviors. Ada seemed to ask redundant questions in the end, (eg, asking about eye pain when this was the primary symptom entered at the beginning). It could be speculated that this functionality serves to add input for the diagnosis to the database. Ada circumvents the common “black box” problem in ML [32] by offering a pictorial description of how many patients in 10 with the given symptoms have the suggested diagnosis. This additional information seems to indicate that by Ada’s accounts, the provided symptoms for dry eyes seemed to correspond less with the diagnosis in 2020 (“8 in 10” vs “5 in 10 people”). Here, one could speculate that there are problems incorporating the data the app accumulates over the years. Generally, the provided statistics indicate that either Bayesian probabilities are used in some way, as artificial neural network (ANN) output activities are not linked to statistical values, or the offered values are interpolated from ANN outputs. The former assumption seems to be backed up by information published by Ada Health itself, where the use of Bayesian networks is mentioned [33]. Buoy asked several questions that seemed off topic (eg, for the user’s health insurance), and in 2018 it presented users with pictures of medical conditions for comparison that may not be suitable for laymen, like testing a patellar reflex in the dry eye case or comparing one’s cornea with a microscopy picture to identify Horner-Trantas dots. In 2020, no pictures were offered for the same set of symptoms. In both years, both Babylon and Buoy failed to produce a useful diagnosis and also gave out very few treatment recommendations, with some results being very far off, such as “Bone issue” or “Non-bacterial brain inflammation” diagnoses by Buoy for the retinal tear patient, contributing to the overall result that Ada and Your.MD fared significantly better in the test than the other two. The variety of treatment recommendations given by the apps for the same starting sets of symptoms is also remarkable. Ada made it simple by generally recommending emergency care for virtually every diagnosis, which may help the manufacturers shift responsibility to the patient but counteracts the possible value of good medical advice. Babylon seemed to send any patient who chooses “severe pain” as a symptom to the emergency department, which is a good outcome for the glaucoma patient, but no diagnosis was given, and its other recommendations were very general. In 2020, Buoy gave the glaucoma patient the advice to seek medical advice within three days as its first option, followed up by “emergency treatment” as second and third options, which would confuse a real patient. Your.MD provided the most valid recommendations in this study, but also worsened on the dry eyes diagnosis from 2018 to 2020, now unnecessarily transitioning from self-treatment to emergency care.

While the number of questions the apps asked did not significantly change between the years, the temporal variances in diagnoses and treatment recommendations indicate the use of learning algorithms in all four, suggesting that the algorithms used for history-taking and diagnosis calculation are changing over time and would thus fall under the FDA’s proposed regulations for learning or “nonlocked” algorithms. In terms of their effectiveness in diagnosing ophthalmic diseases, the results were mixed with a tendency to worsening. It is noteworthy that no trend to improvement of history-taking and results could be observed at all. On the contrary, Ada and Your.MD worsened in their diagnostic outputs, while Babylon and Buoy were stable at a low level. This deterioration of diagnostic performance seems to contradict the very purpose of using “learning” algorithms in the first place and certainly justifies further inquiry. It is also notable that while two of the apps ask more questions than the other two, there seems to be no correlation between the number of questions asked and the quality of the results. On the contrary, the app with the highest overall score had the second-lowest total of questions asked. This indicates high variation in their diagnostic approaches and efficiencies, all worthy of subsequent systematic evaluation. Their algorithms are undisclosed; judging by the apps’ workflows, they all basically resemble the adaptive feedforward neural network–based mobile diagnosis engine that the author conceptualized in 2016 [34], which in its framework resembles the classic AI game “20 Questions” [35]. In both frameworks, two separate neural networks (or similar algorithms) separately calculate the current diagnosis based on the available information and the next best question based on the input up to that point. Where these examples used simple ANNs, the apps may also utilize the previously mentioned RNNs, Bayesian networks, or convolutional neural networks [15], accessed through the chat bot.

There are several limitations to this study. First, the cases that were entered included information that seemed irrelevant to the author for deciding on a suspicion of a potentially dangerous diagnosis; this includes, for example, the prevalence of diabetes, a history of smoking, and seemingly unrelated questions. It is possible, however, that to a large database that collects and sorts information without bias, answering the questions from a biased physician's point of view might in fact mislead the algorithms. Second, the evaluation of the results is as subjective as entering the symptoms, which ultimately might test the apps' ability to imitate a potentially flawed physician rather than whether they can correctly identify diagnoses. This could be improved by introducing systematic evaluations in the framework of a randomized controlled trial. Semigran et al had used human input and output on a randomized stack of diagnoses to assess self-diagnosis apps [7]. New methods may be necessary to investigate AI-driven apps in the future, possibly including some degree of automation considering the superhuman data storage capacities such systems can house and taking into consideration the dynamic of the algorithms. A simpler but also less systematic approach would be to include more physicians in the evaluation of the apps and average their assessments as has been done before [10]. The possibility of the manufacturers adapting to known sets of questions (eg, from this study) should also be considered upon further investigation. Third, the sample size is low. In future investigations, large-sample investigations should be preferred. Other authors like Fraser et al in 2018 have already demanded standardized and transparent procedures for examining such devices [36]. In 2019, Kelly et al advocated for a focus on peer-reviewed studies in order to increase trust in AI devices and added that the introduction of consumer-oriented technology offers the opportunity for vast prospective studies with the new collected data, provided that a sufficient level of data transparency is reached [37]. They also mentioned that an extension to the existing TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement, which has defined recommendations for the evaluation of diagnosis prediction since 2015 [38], is being developed to include ML algorithms [39]. In addition, the World Health Organization and the International Telecommunication Union are working on benchmarking frameworks for AI tools in health care [40]. All of these could act as guidelines for future scientific exploration of the topic, but they will require funding and manpower. The four enterprises mentioned in this paper are employing physicians, mostly with an additional formal or informal education in medical informatics or similar, at high ranks within their hierarchy, in two of them even as cofounders [41-44]. Now that this Pandora's box of AI in the hands of patients and corporations has been opened, the question is whether this novel type of physician that supervises and administrates an automated diagnostic system will be mirrored by scientific counterparts who publicly evaluate the apps' performances, or whether these essential data will remain undisclosed—a common practice in the commercial sector due to conflicts of interest. Considering the apps' possible leverage and impact on public health [4], this should be in the public interest. The prospect of AI support for physicians provided by simple and accessible apps in the hands of layman users could be a golden one, as long as they actually learn and improve. Most importantly, they need to satisfy the crucial premises within the field of health care: to be efficient and safe.

Authors' Contributions

The author AC is currently not affiliated with any institution, but is an Independent Scholar.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Walkthrough through all apps and diagnoses.
[DOCX File , 35 KB - jmir_v22i12e18097_app1.docx]

Multimedia Appendix 2
Additional results tables with scores.
[DOCX File , 18 KB - jmir_v22i12e18097_app2.docx]

Multimedia Appendix 3
Additional tables (no. of questions asked, time taken).
[DOCX File , 15 KB - jmir_v22i12e18097_app3.docx]

References

1. Willson M. Algorithms (and the) everyday. Inf Commun Soc 2016 Jun 20;20(1):137-150. [doi: 10.1080/1369118X.2016.1200645]
2. Makridakis S. The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. Futures 2017 Jun;90:46-60. [doi: 10.1016/j.futures.2017.03.006]

3. Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S. The Use of Artificially Intelligent Self-Diagnosing Digital Platforms by the General Public: Scoping Review. JMIR Med Inform 2019 May 01;7(2):e13445 [FREE Full text] [doi: 10.2196/13445] [Medline: 31042151]
4. Ćirković A. AI in Self-Diagnosis - History, Theoretical Foundations, Potentials and Current Status. ResearchGate 2021:6-8 Preprint(forthcoming) [FREE Full text] [doi: 10.13140/RG.2.2.33461.83684]
5. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, et al. Current Applications and Future Impact of Machine Learning in Radiology. Radiology 2018 Aug;288(2):318-328 [FREE Full text] [doi: 10.1148/radiol.2018171820] [Medline: 29944078]
6. Standing S, Standing C. Mobile technology and healthcare: the adoption issues and systemic problems. Int J Electron Healthc 2008;4(3-4):221-235. [doi: 10.1504/IJEH.2008.022661] [Medline: 19174359]
7. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015;351:h3480 [FREE Full text] [Medline: 26157077]
8. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a Chatbot (Ada) in the Diagnosis of Mental Disorders: Comparative Case Study With Lay and Expert Users. JMIR Form Res 2019 Oct 29;3(4):e13863 [FREE Full text] [doi: 10.2196/13863] [Medline: 31663858]
9. Middleton K, Butt M, Hammerla N, Hamblin S, Mehta K, Parsa A. Sorting out symptoms: design and evaluation of the 'babylon check' automated triage system. arXiv e-prints. 2016 Jun 7. URL: http://arxiv.org/abs/1606.02041 [accessed 2020-01-16]
10. Razzaki S, Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, et al. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. arXiv e-prints. 2018 Jun 27. URL: http://arxiv.org/abs/1806.10698 [accessed 2019-01-17]
11. United States Food and Drug Administration. Examples of Software Functions for Which the FDA Will Exercise Enforcement Discretion. FDA. 2019 Sep 26. URL: https://tinyurl.com/y243bah9 [accessed 2018-04-24]
12. United States Food and Drug Administration. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback. 2019. URL: https://www.fda.gov/media/122535/download [accessed 2020-01-03]
13. Garbuio M, Lin N. Artificial Intelligence as a Growth Engine for Health Care Startups: Emerging Business Models. Calif Manage Rev 2018 Nov 21;61(2):59-83. [doi: 10.1177/0008125618811931]
14. Winn AN, Somai M, Fergestrom N, Crotty BH. Association of Use of Online Symptom Checkers With Patients' Plans for Seeking Care. JAMA Netw Open 2019 Dec 02;2(12):e1918561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18561] [Medline: 31880791]
15. Yao C, Qu Y, Jin B, Guo L, Li C, Cui W, et al. A Convolutional Neural Network Model for Online Medical Guidance. IEEE Access 2016;4:4094-4103. [doi: 10.1109/access.2016.2594839]
16. Buhr S. TechCrunch. Buoy hopes to fight fake online health news with an artificially intelligent app. 2017 Mar 08. URL: http://social.techcrunch.com/2017/03/08/buoy-aims-to-fight-fake-online-health-news-with-an-artificially-intelligent-app/ [accessed 2020-01-31]
17. Khondkaryan A, Francis BA. Angle-Closure Glaucoma. American Academy of Ophthalmology. 2013 Dec 18. URL: https://www.aao.org/munnerlyn-laser-surgery-center/angleclosure-glaucoma-19 [accessed 2018-12-20]
18. Prophylactic Treatment of Retinal Breaks. In: 2020-2021 Basic and Clinical Science Course (BCSC), Section 02: Fundamentals and Principles of Ophthalmology. San Francisco: American Academy of Ophthalmology; 2020.
19. Akpek EK, Amescua G, Farid M, Garcia-Ferrer FJ, Lin A, Rhee MK, American Academy of Ophthalmology Preferred Practice Pattern Cornea External Disease Panel. Dry Eye Syndrome Preferred Practice Pattern®. Ophthalmology 2019 Jan;126(1):P286-P334. [doi: 10.1016/j.ophtha.2018.10.023] [Medline: 30366798]
20. American Academy of Ophthalmology. Basic and Clinical Science Course Complete Set 2012-2013. San Francisco: American Academy of Ophthalmology; 2012.
21. Leigh L. Europe-founded, New Zealand-trialled app Ada offers a virtual assessment of health symptoms in real time. Startup Daily. 2016 Aug 11. URL: http://www.startupdaily.net/2016/08/europe-founded-new-zealand-trialled-app-ada-offers-virtual-assessment-health-symptoms-real-time/ [accessed 2018-04-14]
22. New 'smart' health app Ada hits No.1 spot in New Zealand. Scoop Independent News. 2016 Aug 11. URL: http://www.scoop.co.nz/stories/SC1608/S00037/new-smart-health-app-ada-hits-no1-spot-in-new-zealand.htm [accessed 2018-04-14]
23. Solon O. Babylon app puts a GP in your pocket. Wired. 2014 Apr 28. URL: https://www.wired.co.uk/article/babylon-ali-parsa [accessed 2018-04-18]
24. Senior Data Science Engineer. AngelList. 2020. URL: https://angel.co/company/babylon_health/jobs/955495-senior-data-science-engineer [accessed 2020-11-25]
25. Ni L, Lu C, Liu N, Liu J. MANDY: Towards a Smart Primary Care Chatbot Application. Singapore: Springer; 2017 Presented at: Knowledge and Systems Sciences: 18th International Symposium, KSS 2017; November 17–19, 2017; Bangkok, Thailand p. 38-52. [doi: 10.1007/978-981-10-6989-5_4]

26. Mack H. Digital health startup Buoy launches AI-powered, symptom-checking chatbot. mobihealthnews. 2017 Mar 08. URL: http://www.mobihealthnews.com/content/digital-health-startup-buoy-launches-ai-powered-symptom-checking-chatbot [accessed 2018-04-14]
27. Buoy Health - A chatbot that helps diagnose your symptoms. producthunt.com. 2017 Mar 12. URL: https://www.producthunt.com/posts/buoy-health [accessed 2018-04-11]
28. Your.MD. Crunchbase. 2018. URL: https://www.crunchbase.com/organization/your-md [accessed 2018-03-23]
29. Your.MD. StartUs. 2018. URL: https://www.startus.cc/company/yourmd [accessed 2018-04-18]
30. Thomas K. Will mobile health apps make GPs redundant? The Guardian. 2016 Apr 16. URL: http://www.theguardian.com/sustainable-business/2016/apr/16/mobile-health-apps-gps-nhs-doctors [accessed 2018-04-18]
31. Rynerson JM, Perry HD. DEBS - a unification theory for dry eye and blepharitis. Clin Ophthalmol 2016;10:2455-2467 [FREE Full text] [doi: 10.2147/OPTH.S114674] [Medline: 28003734]
32. Hart A, Wyatt J. Evaluating black-boxes as medical decision aids: issues arising from a study of neural networks. Med Inform (Lond) 1990;15(3):229-236. [doi: 10.3109/14639239009025270] [Medline: 2232958]
33. Zimmer V. ada \inside. Digital Health Connect. 2018 Jun. URL: https://www.digitalhealthconnect.ch/wp-content/uploads/2018/06/AdaHealth-Vincent-Zimmer_DHC18.pdf [accessed 2020-11-25]
34. Ćirković A. Diagnoseapp als künstliches neuronales Netz [A Diagnosis App Based on Two Artificial Neural Networks]. Berlin: Beuth-University of Applied Sciences; Mar 06, 2016.
35. Burgener R. Artificial neural network guessing method and game. Google Patents. 2006 Oct 12. URL: http://www.google.com/patents/US20060230008 [accessed 2018-10-01]
36. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet 2018 Dec 24;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]
37. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
38. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 07;350:g7594. [doi: 10.1136/bmj.g7594] [Medline: 25569120]
39. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet 2019 Apr 20;393(10181):1577-1579. [doi: 10.1016/S0140-6736(19)30037-6] [Medline: 31007185]
40. Wiegand T, Krishnamurthy R, Kuglitsch M, Lee N, Pujari S, Salathé M, et al. WHO and ITU establish benchmarking process for artificial intelligence in health. Lancet 2019 Jul 06;394(10192):9-11. [doi: 10.1016/S0140-6736(19)30762-7] [Medline: 30935732]
41. Rebecca Sells MBChB (Hons) DSFRH. LinkedIn. URL: https://www.linkedin.com/in/dr-rebecca-sells-mbchb-hons-dsfrh-84b31bb9/ [accessed 2018-04-18]
42. Claire Novorol. LinkedIn. URL: https://www.linkedin.com/in/clairenovorol/?locale=de_DE [accessed 2018-04-18]
43. Andrew Le, MD. LinkedIn. URL: https://www.linkedin.com/in/andrewle21/ [accessed 2018-04-18]
44. András Meczner. LinkedIn. URL: https://www.linkedin.com/in/andr%C3%A1s-meczner-2399ab52/ [accessed 2018-04-18]

Abbreviations

AI: artificial intelligence
ANN: artificial neural network
FDA: US Food and Drug Administration
ML: machine learning
NLP: natural language processing
RNN: recurrent neural network
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

---

Edited by G Eysenbach; submitted 06.02.20; peer-reviewed by A Palanica, D Mendes, J Xu; comments to author 10.03.20; revised version received 04.08.20; accepted 30.10.20; published 04.12.20.

Please cite as:
Ćirković A
Evaluation of Four Artificial Intelligence–Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study
J Med Internet Res 2020;22(12):e18097
URL: https://www.jmir.org/2020/12/e18097
doi:10.2196/18097
PMID:33275113

Review

Artificial Intelligence in the Fight Against COVID-19: Scoping Review

Alaa Abd-Alrazaq[1], PhD; Mohannad Alajlani[2], PhD; Dari Alhuwail[3], PhD; Jens Schneider[1], PhD; Saif Al-Kuwari[1], PhD; Zubair Shah[1], PhD; Mounir Hamdi[1], PhD; Mowafa Househ[1], PhD

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

[2]Institute of Digital Healthcare, University of Warwick, Coventry, United Kingdom

[3]Information Science Department, College of Life Sciences, Kuwait University, Kuwait, Kuwait

Corresponding Author:
Mowafa Househ, PhD
Division of Information and Computing Technology, College of Science and Engineering
Hamad Bin Khalifa University
Qatar Foundation
PO Box 5825, Doha Al Luqta St, Ar-Rayyan
Doha,
Qatar
Phone: 974 55708549
Email: mhouseh@hbku.edu.qa

Abstract

Background: In December 2019, COVID-19 broke out in Wuhan, China, leading to national and international disruptions in health care, business, education, transportation, and nearly every aspect of our daily lives. Artificial intelligence (AI) has been leveraged amid the COVID-19 pandemic; however, little is known about its use for supporting public health efforts.

Objective: This scoping review aims to explore how AI technology is being used during the COVID-19 pandemic, as reported in the literature. Thus, it is the first review that describes and summarizes features of the identified AI techniques and data sets used for their development and validation.

Methods: A scoping review was conducted following the guidelines of PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews). We searched the most commonly used electronic databases (eg, MEDLINE, EMBASE, and PsycInfo) between April 10 and 12, 2020. These terms were selected based on the target intervention (ie, AI) and the target disease (ie, COVID-19). Two reviewers independently conducted study selection and data extraction. A narrative approach was used to synthesize the extracted data.

Results: We considered 82 studies out of the 435 retrieved studies. The most common use of AI was diagnosing COVID-19 cases based on various indicators. AI was also employed in drug and vaccine discovery or repurposing and for assessing their safety. Further, the included studies used AI for forecasting the epidemic development of COVID-19 and predicting its potential hosts and reservoirs. Researchers used AI for patient outcome–related tasks such as assessing the severity of COVID-19, predicting mortality risk, its associated factors, and the length of hospital stay. AI was used for infodemiology to raise awareness to use water, sanitation, and hygiene. The most prominent AI technique used was convolutional neural network, followed by support vector machine.

Conclusions: The included studies showed that AI has the potential to fight against COVID-19. However, many of the proposed methods are not yet clinically accepted. Thus, the most rewarding research will be on methods promising value beyond COVID-19. More efforts are needed for developing standardized reporting protocols or guidelines for studies on AI.

(J Med Internet Res 2020;22(12):e20756) doi:10.2196/20756

KEYWORDS

artificial intelligence; machine learning; deep learning; natural language processing; coronavirus; COVID-19; 2019-nCoV; SARS-CoV-2

Introduction

Background

COVID-19 broke out in Wuhan, Hubei Province, China in December 2019 [1], spreading across the world and, as of May 2020, claiming the lives of more than 330,000 people [2]. Caused by SARS-CoV-2, COVID-19 was declared a global pandemic by the World Health Organization in March 2020 [3]. Many individuals infected with COVID-19 experienced fever, dry cough, and fatigue; some faced a severe course of the medical condition, often requiring intensive care, including mechanical ventilation [4]. The contagious COVID-19 and its unprecedented volume of cases around the world have caused national and international disruptions to business, health care, education, transportation, and nearly every aspect of our daily lives [5]. Prompt and effective countermeasures are necessary to cap off the effects of this pandemic; comprehensive public health strategies that involve surveillance, diagnostics, clinical treatment, and research are required [6].

Leveraging digital tools and technologies to combat COVID-19 can augment public health strategies [7], for example, by using chatbots to address public inquiries about COVID-19. Additionally, using digital tools, public health professionals can track in real time the incidence of COVID-19 infections and potentially model their projection. Among such tools is artificial intelligence (AI)—a branch of computer science concerned with intelligently analyzing and handling complex information [8,9]—amplifying public health efforts against COVID-19. Despite the enthusiasm for AI applications since the 1950s, only recently have we witnessed interest in AI due to the availability of high-performance computing and vast amounts of data being generated every second [10].

AI enables machines to become intelligent, understand queries, sift through and connect mountains of data points, and draw actionable conclusions [11]. Although defining the taxonomy of AI is not trivial, its methods can be categorized based on the objective pursued: learn from knowledge, explore and discover knowledge, extract conclusions, and reason from knowledge [8].

Soon after the COVID-19 pandemic spread across the world, several governments, research institutes, and technology companies have issued calls to action urging researchers to develop AI applications to assist with COVID-19–related research [12]. From a hierarchical perspective, AI can support COVID-19 at different levels: the molecular level (eg, drug and vaccine discovery), patient level (eg, patient diagnosis), and population level (eg, epidemiological surveillance) [13].

A full review of the AI field is beyond the scope of this review, and we refer the reader to some surveys (eg, [14]) and lectures (eg, [15,16]). However, we provide a compact overview of the AI-based techniques occurring most frequently in included studies in Multimedia Appendix 1.

Research Problem

AI has the ability to analyze big data sets through aggregating and sifting through mountains of health care data (including patient data) to generate insights that can enable predictive analysis. The quick ability to obtain these insights helps clinicians as well as other stakeholders in the health care ecosystem to make effective, safe, and timely decisions to better serve patients and public health policy. There has been a steady rise in the number of studies regarding the use of AI techniques to resolve or address the COVID-19 pandemic [13]. Much of the AI research effort during the COVID-19 pandemic has been scattered, and a need to explore and summarize how AI technologies are being used to resolve or address the many challenges relating to COVID-19 can help us plan on how to leverage AI technologies in the current or a future pandemic. Several reviews have been conducted on AI techniques used to address the COVID-19 pandemic [12,13,17-20]. However, much of the work has been in the form of literature reviews [12,13,17-19] or systematic reviews focusing on one application of AI (eg, diagnosis and prognosis of COVID-19) [20]. Therefore, it is necessary to conduct a more systematic and comprehensive review that focuses on all applications of AI used amid the COVID-19 pandemic. Accordingly, this review aims to explore how AI technology is being used during the COVID-19 pandemic as reported in the literature. The results can be useful for health care professionals and policy makers considering leveraging AI to complement public health efforts in response to COVID-19.

Methods

To achieve the objective of this study while ensuring both replicable and transparent methods, we conducted a scoping review following the guidelines of PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [21]. Methods used in this review are detailed in the following subsections.

Search Strategy

Search Sources

In this review, we performed search queries between April 10 and 12, 2020, on the following online databases: MEDLINE (via Ovid), EMBASE (via Ovid), PsycInfo (via Ovid), IEEE Xplore, ACM Digital Library, arXiv, medRxiv, bioRxiv, Scopus, and Google Scholar. In the case of Google Scholar and due to the volume of returned hits, only the first 100 results were considered, as we found that, beyond this, results quickly lose relevance and applicability. In addition to searching bibliographic databases, we screened the reference list of the included studies and relevant reviews to look for other relevant studies that could be added to this review (ie, backward reference list checking).

Search Terms

The search terms we used to identify relevant studies were specified from the available literature and by referring to subject matter experts. These terms were selected based on the target intervention (eg, AI, machine learning, and deep learning) and the target disease (eg, coronavirus, COVID-19, and 2019-nCoV). Details about the exact search strings used in this study are provided in Multimedia Appendix 2.

Study Eligibility Criteria

In this review, we focused on any AI-based technology or approach used for any purpose related to the COVID-19 pandemic, such as diagnosis, epidemiological predictions, treatment and vaccine discovery, and prediction of patient outcomes. However, we excluded studies providing an overview or proposing a potential AI technique for COVID-19, or studies that were purely discussed from a research perspective.

We considered studies published in English between December 25, 2019, and April 12, 2020, such as peer-reviewed articles, theses, dissertations, conference proceedings, and preprints, while excluding other publications such as reviews, conference abstracts, proposals, editorials, and commentaries. We did not enforce any restrictions on the country of publication, study design, comparator, and outcomes.

Study Selection

Two reviewers, namely, authors AAA and MA, independently screened the titles and abstracts of the identified studies. They independently read the full text of studies that passed the *title and abstract* screening. We then investigated any disagreement between AAA and MA and resolved them through discussion and consensus. We calculated Cohen kappa [22] to measure the reviewer's agreement and found it to be 0.83 for the *title and abstract* screening and 0.94 for the full-text reading, indicating a very good agreement [23]. Multimedia Appendix 3 shows a matrix of interrater agreement in each step.

Data Extraction

Multimedia Appendix 4 shows a purpose-built data extraction form, which was pilot-tested using 7 relevant studies to accurately extract data. The two reviewers (AAA and MA) independently extracted data related to characteristics of the included studies, AI techniques, and data sets used for the development and validation of AI models. Like the study selection process, any disagreement between the reviewers was resolved through consensus. We calculated Cohen kappa [22] and found it to be 0.88, meaning a very good agreement [23].

Data Synthesis

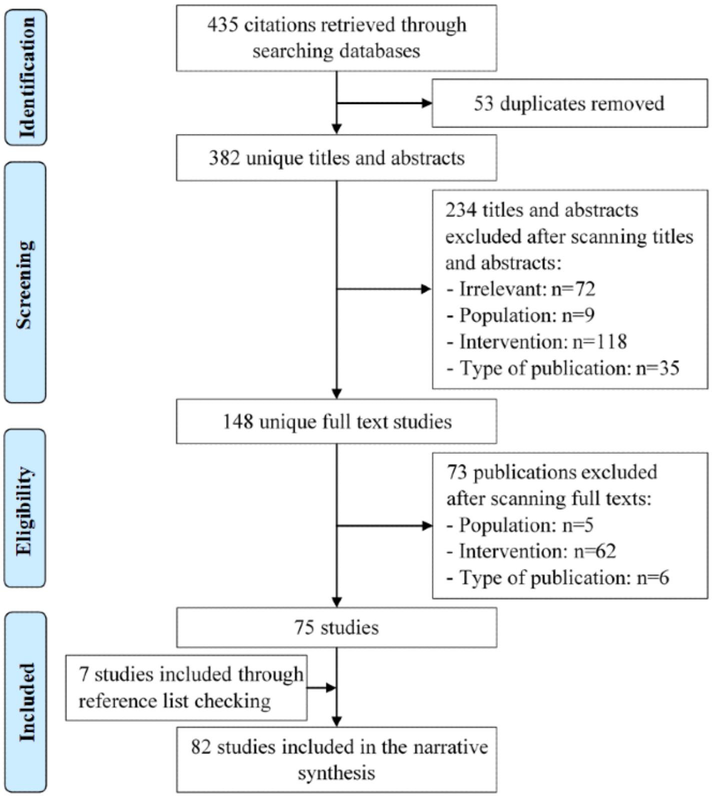
After extracting the data from the identified studies, we used a narrative approach to synthesize it. Specifically, we classified and described AI techniques used in the included studies in terms of their purposes (eg, diagnosis and drug and vaccine development), AI area or branch (eg, traditional machine learning and deep learning), AI models and algorithms (eg, decision tree, random forest, and naive Bayes), and platform (ie, computer and mobile). Further, we described the data sets used for development and validation of AI models in terms of data sources (eg, public databases and clinical settings); type of data (eg, radiology images, biological data, and laboratory data); size of the data set; type of validation; and proportion of training, validation, and test data sets. We used Microsoft Excel (Microsoft Corporation) to manage data synthesis.

Results

Search Results

We retrieved 435 studies through searching the identified bibliographic databases (Figure 1). Of those studies, we removed 53 duplicates; we then screened the titles and abstracts of the remaining 382 studies. The screening process led to the exclusion of 234 studies for reasons detailed in Figure 1. After reading the full texts of the remaining 148 studies, we excluded 73 studies, as they did not meet all eligibility criteria. Thus, we included the remaining 75 studies. We identified 7 additional studies by checking reference lists of the included studies and relevant literature reviews. Overall, 82 studies were included in this review.

Figure 1. Flowchart of the study selection process.



Characteristics of the Included Studies

Among the included studies, 72 were preprints and 10 were published articles in peer-reviewed journals (Table 1 and Figure 2). About two-thirds (n=53) of the included studies were submitted in March 2020, and the remaining studies were published in February and April 2020. However, no studies were published during the first 2 months of the COVID-19 outbreak. The included studies were conducted in 19 countries; however, half of the studies (n=41) were published in China. Multimedia Appendix 5 shows the characteristics of each included study.

Table 1. Characteristics of the included studies.

Characteristics	Studies (N=82), n
Paper status	
Preprint	72
Published	10
Submission month	
February	13
March	53
April	16
Country of publication	
China	41
US	9
India	6
Turkey	5
Canada	4
UK	3
Bangladesh	2
Austria	1
Egypt	1
Greece	1
Hong Kong	1
Hungary	1
Japan	1
Korea	1
Netherlands	1
Pakistan	1
Qatar	1
Sudan	1
Switzerland	1

Figure 2. Publications by months and country.

AI-Based Techniques Used for COVID-19

Purposes or Uses of AI Against COVID-19

As shown in Table 2, AI techniques have been used for five purposes amid the pandemic. In 31 studies [24-54], AI was used for diagnosing COVID-19 cases or identifying suspected COVID-19 cases based on various indicators, including computed tomography (CT) images [24-38], x-ray images [39-50], laboratory tests [51,52], genome sequences [53], and respiratory patterns [54].

Table 2. Purposes and uses of artificial intelligence against COVID-19.

Purposes/uses	Studies (N=82), n
Diagnosis	
CT[a] images	15
X-ray images	12
Laboratory tests	2
Genome sequence	1
Respiratory patterns	1
Treatment and vaccines	
Drug discovery	9
Vaccine discovery	4
Protein structure	4
Drug repurposing	2
Treatment safety	1
Epidemiology	
Epidemic development	14
Potential reservoirs	3
Patient outcome	
Severity	6
Progression to severe	4
Mortality risk	2
Risk factors	1
Hospital stay	1
Infodemiology	
Raising awareness	1

[a] CT: computed tomography.

In 20 studies [55-74], AI was also harnessed for treatment and vaccines for COVID-19. Specifically, 9 studies used AI for discovering drugs suitable for combating COVID-19 [55-63], and 2 studies used AI for repurposing commercially available drugs that could be used for treating COVID-19 [64,65]. There was 1 study that employed AI to predict the safety of using traditional Chinese medicine for COVID-19 [66]. In 4 studies [67-70], AI was used for discovering COVID-19 vaccines. Another 4 studies used AI for predicting the protein structure of SARS-CoV-2, thereby aiding researchers and pharmaceutical companies to discover drugs for COVID-19 [71-74].

There were 17 studies that used AI for epidemiological modeling tasks [75-91]. In particular, 14 of these studies employed AI for forecasting the epidemic development (eg, numbers of confirmed, recovered, death, suspected, asymptomatic, and critical cases, and lengths and ending time) [75-88], and 3 studies [89-91] used AI for predicting the potential hosts or reservoirs of SARS-CoV-2.

In 14 studies [33,92-104], AI was used for patient outcome–related tasks. In particular, 6 studies used AI for segmentation and quantification of infected regions in the lungs due to COVID-19, thereby assessing the severity of the disease [92-97]. AI was also used in 4 studies for identifying cases at high risk of progression to severe COVID-19 [33,98-100]. Furthermore, AI was also used for predicting mortality risk [101,102], its associated factors [103], and the length of hospital stay in patients with COVID-19 [104].

AI has also been used for infodemiology [105]. Specifically, AI was used for raising awareness to use water, sanitation, and hygiene through combining authentic sources of information with daily news [105]. Multimedia Appendix 5 presents the purposes or uses of AI techniques in each included study.

Features of AI-Based Techniques Used for COVID-19

In 29 studies [24,31,47, 51,52,58, 63,68,70, 74,79,83-86,88,90,91,95-105], AI techniques used against COVID-19 were based on traditional machine learning models and algorithms (Table 3). The most commonly used machine learning models and algorithms were support vector machine (SVM) [24, 31, 47,58, 68, 70,79,91,98,101], random forest [31,58,68,74,90,96,101,103,104], decision tree [52,58,68,74,79,97,99,101,102], and logistic regression [31,51,52,58,68,99-101,104].

Table 3. Features of AI-based techniques used for COVID-19.

Features	Studies (N=82), n
AI[a] branches[b]	
Deep learning	60
Machine learning	29
Natural language processing	3
AI models/ algorithms[c]	
Convolutional neural network	37
Support vector machine	10
Random forest	9
Decision tree	9
Logistic regression	9
Recurrent neural network	8
Artificial neural network (unspecified)	6
Transfer learning	4
Autoencoders	4
Deep neural network	3
K-nearest neighbors	3
Least absolute shrinkage and selection operator	3
Polynomial neural network	3
Multilayer perceptron	2
Advance deep Q-learning network	2
AdaBoost	1
Auto-regressive integrated moving average model	1
Bayesian analysis	1
Bidirectional encoder representations from transformers	1
Continuous bag of words	1
Eureqa modeling	1
Genetic algorithm	1
Generative adversarial network	1
Generalized logistic growth model	1
Holistic agent-based model	1
Linear discriminant analysis	1
Linear regression	1
Language model	1
Multi-task deep model	1
Naive Bayes	1
Porter stemming	1
Reinforcement learning	1
Skip-gram model	1
Time series forecasting	1
Universal-sentence-encoder-large	1
Vector auto average	1

Features	Studies (N=82), n
Platforms	
Computer	81
Mobile	1

[a] AI: artificial intelligence.

[b] Numbers do not add up as AI techniques in some studies were based on more than one AI branch.

[c] Numbers do not add up as several studies used more than one AI model or algorithm.

In 60 studies, AI techniques used against COVID-19 were based on deep learning models and algorithms [25-50,53-57,59-67,69,71-73,75-82,87,89,92-95,98,101]. The most commonly used learning models and algorithms in the included studies were convolutional neural network (CNN) [25-50,53,62,64,72,73,82,89,92-95] and recurrent neural network (RNN) [54,55,57,59,71,73,77,98].

In 2 studies [64,105], AI techniques used against COVID-19 were based on models related to natural language processing (NLP), such as the continuous bag of words model, skip-gram models, and porter stemming. Although AI techniques were implemented in mobile phones for 1 study [105], computers were the platform for AI techniques in the remaining studies. Multimedia Appendix 5 shows features of AI-based techniques used in each included study.

Features of Data Sets Used for Development and Validation of AI Models

As shown in Table 4, public resources (eg, National Center for Biotechnology Information [NCBI], GitHub, and Kaggle) were the most commonly used data source for development and validation of AI models [24,27, 29,36,39-50,53, 55-65,67-75,77, 80-85,87-89,91-93,103,105]. Other data sources used by the included studies were as follows: clinical settings (eg, databases in hospitals and medical centers) [25-35,37,38,51,52,63,94-98,100, 102,104], government sources (eg, Chinese Center for Disease Control and Prevention) [53,76,78, 79, 84,86, 90, 99,101], literature (eg, previous studies and books) [36,40,42,61,66,101], news websites [101,105], and participants recruited by the study [54].

Table 4. Features of data sets used for development and validation of artificial intelligence models.

Features	Studies (N=82), n
Data sources[a]	
Public databases	52
Clinical settings	24
Government sources	9
Literature	6
News websites	2
Participants	1
Data types[b]	
Radiology image	35
Biological data	23
Epidemiological data	15
Clinical data	11
Laboratory data	8
Demographic data	5
Guidelines	1
News articles	1
Data set size[c]	
<1000	26
1000-9999	16
≥10,000	8
Type of validation[d,e]	
Train-test split	25
K-fold cross-validation	18
External validation	11
Proportion of training set (%)[f]	
≤25	3
26-50	2
51-75	16
>75	28
Proportion of validation set (%)[g]	
≤25	8
26-50	3
51-75	0
>75	0
Proportion of test set (%)[h]	
≤25	35
26-50	10
51-75	3
>75	1

[a] Numbers do not add up as several studies collected their data from more than one data source.

[b] Numbers do not add up as several studies collected more than one type of data.

[c] Data set size was reported in 50 studies.

[d] Type of validation was reported in 53 studies.

[e] Numbers do not add up as 1 study used two different types of validation.

[f] Proportion of the training set was reported in 49 studies.

[g] Proportion of the validation set was reported in 11 studies.

[h] Proportion of the test set was reported in 49 studies.

The types of data collected from these data sources were as follows: radiology images (eg, CT and x-ray) [24-50,54,92-96,98,104], biological data (eg, protein and genome sequences) [53,55-65,67-74,89-91], epidemiological data (eg, number of infected and recovered cases) [75-85,87,88,97,102], clinical data (eg, signs, symptoms, physician notes, and patients' history) [25,51,52,66,97-103], laboratory data (blood and polymerase chain reaction test results) [25,51,52,86,97,98,100,102], demographic data (eg, age, gender, and ethnicity) [52,99-102], guidelines [105], and news articles [105].

The data set size was reported by 50 studies, ranging from 31 to 3,000,000. The data set size was less than 1000 samples in half of these studies [24,27,32,34, 36,37,39, 41,44,45, 47,51-53,69,86,92,94-98,100,102-104], and only 8 studies reported a size of 10,000 samples or more [25,26,54,59,61,87,99,101].

Validation of models was reported in 53 studies. Three types of validation were used in the included studies: train-test split [25,29, 30, 34-39, 41, 43, 44, 47-50, 59, 75,87,88, 93,94,99,100,103], K-fold cross-validation [24,31,40, 42,45,46, 52,53, 58, 66,68,90-92,96,98, 101,104], and external validation [26,27,29,32,33,38,51,54,82,95,102].

The training set proportion of the total data set was reported in 49 studies. The proportion of the training set ranged from ≤25% in 3 studies [25,27,28], 26%-50% in 2 studies [26,95], 51%-75% in 16 studies [32,33,35,36, 38, 39,47, 51, 59,75,87,88,96,100,102,103], and >75% in 28 studies [24,29-31,34,37,40-46,48,50, 52-54,58, 66,90-92, 94,98,99,101,104]. The mean of the proportions of the training set in the 49 studies was 72.7%.

The validation set proportion of the total data set was reported in 11 studies; it ranged from ≤25% in 8 studies [26,28, 35, 36,38, 47, 48,53] and 26%-50% in 3 studies [25,100,102]. The mean of the validation set proportions in the 11 studies was 18.7%.

The test set proportion of the total data set was reported in 49 studies, ranging from ≤25% in 35 studies [24,29-31,34,36,37,40-48,50,52-54,58,59,66,75,87,90-92,94,98-102,104], 26%-50% in 10 studies [26,32,33,35,38,39,51,88,96,103], 51%-75% in 3 studies [25,28,95], and >75% in 1 study [27]. The mean of the test set proportions in the 49 studies was 22.9%. Multimedia Appendix 6 presents features of the data sets used for development and validation of AI models in each included study.

Discussion

Principal Results

In this study, we conducted a scoping review of the use of AI against COVID-19. We found a lack of publications in December 2019 and January 2020. This is not surprising, given that SARS-CoV-2 was only identified on January 7 [106]; insufficient data was not available to back scientific publications, in particular internationally; and the contagiousness and aggressiveness of the virus were underestimated (first lockdown in China was January 23 [106]). Half of the studies in this report were published in China. Since SARS-CoV-2 originated in China and affected it the most during the first 3 months of the pandemic, it had the most data related to COVID-19. Considering lengthy publication processes and the vast number of COVID-19–related manuscript submissions, it is also not surprising that most of the included studies were preprints.

In the included studies, AI was used for five purposes: diagnosis, treatment and vaccine discovery, epidemiological modeling, patient outcome–related tasks, and infodemiology. None of the included studies used AI for other purposes such as contact tracing of the individuals, providing training to students and health care professionals, or robotics to deal with suspected and quarantined cases.

Most of the AI techniques used in the included studies were based on deep learning approaches such as CNN and RNN. All but 1 study used desktop machines, workstations, and clusters as opposed to mobile platforms. This can be explained by the computational demand in training AIs. Although all major mobile phone manufacturers equip their flagship models with AI coprocessors, these coprocessors accelerate inference, a computationally much lighter task. In addition, *federated* learning [107] (a machine learning privacy-preserving technique usually used in mobile phones) is still in its infancy and raises issues such as data sovereignty, scalability, and performance.

Data sources used in the included studies usually came from the public domain (eg, NCBI, GitHub, Kaggle) and proprietary databases (less common). Radiology images were the most commonly used type of data, followed by biological data. The number of samples was still comparably small (less than 1000 in half of the studies). The diversity and size of data indicate a lack of publicly available data despite COVID-19 cases having surpassed 5 million at the time of writing. We, therefore, second Wynants et al [20] call "for immediate sharing of the individual participant data from COVID-19 studies worldwide."

Practical and Research Implications

Although this review explores the use of AI against COVID-19, some applications could prove useful far beyond this pandemic. For instance, Kiwibot designs autonomous medical delivery robots to minimize interpersonal contact [108]. Whiteboard Coordinator developed a high precision thermal screening device eliminating individual measurements, leading to higher throughput and larger social distances [109]. Although mobile phones are not yet the AI platform of choice, the first apps to track interpersonal contact using mobile phones have been published to prioritize COVID-19 testing [110]. Finally, whereas a real-time reverse transcription polymerase chain reaction test takes around 25 minutes and requires stocks of chemical reagents, AI can inspect chest CTs to provide preliminary diagnoses in seconds. We believe that increasing social distance and providing fully autonomous checkups will be the most valuable use of AI beyond the current pandemic.

In the past, fundamental AI research was focused mainly on faster (or even feasible) training. We believe that, in the future, this must be complemented with public education. AI mistrust, because of our still lacking understanding of how AI works at the deepest level, further raises ethical questions that need to be answered before AI will be uniformly accepted. We also found that AI features and results were reported in an inconsistent manner, potentially fueling AI mistrust and making a direct comparison between studies difficult. Of the 82 studies, we found that only 64.6% (n=53) of the studies included in this review disclosed the type of validation, 61% (n=50) mentioned the data size, and more than 7% (n=6) did not even specify the type of AI used. It is therefore important that we as a community develop a standardized reporting protocol to slow down the barrage of poorly conducted COVID-19 studies that threaten to overwhelm serious scientists (1916 related papers were retrieved before April 5, 2020 by Wynants et al [20]), strengthen properly conducted studies, and improve reproducibility.

We found that, explicably, the landscape of studies is still dominated by Chinese institutions, which bears the potential for cultural, technological, and geospatial biases. However, we see a recent move toward a more balanced landscape (see Figure 2). Although we identified more than 100 models developed in the included studies, we did not assess their quality as it was out of this review's scope. Therefore, further reviews are needed to assess the quality of AI models used in the fight against COVID-19.

Given the current "infodemic" [13], we find it surprising that NLP is not used more often. We see AI-based analysis of effective advertisement for nonpharmaceutical interventions as one research opportunity to answer questions like what manner of speech results in maximum public acceptance.

Strengths and Limitations

Strengths

Given that this review includes all AI techniques used for the COVID-19 pandemic regardless of their characteristics, study design, study setting, and country of publication, it may be considered the most comprehensive review in this research area. This helps readers to speculate how AI is being leveraged amid the COVID-19 pandemic. In comparison with similar reviews [12,13,17-20], our review is the only one that describes and summarizes features of the identified AI techniques and data sets used for their development and validation. Furthermore, unlike previous reviews [12,13,17-19], this review follows the full scientific rigor of PRISMA-ScR [21].

In contrast to other reviews, we searched the most commonly used databases in health and information technology fields to identify as many relevant studies as possible. Thus, the number of studies included in this review was much higher than in other reviews [12,13,17-20]. Additionally, we strove to retrieve gray literature and minimized the risk of publication bias by searching Google Scholar and conducting backward reference list checking. Furthermore, we minimized selection bias by having two independent reviewers conducting study selection and data extraction, with a very high agreement in both processes.

Limitations

Given that our review excludes proposals of AI techniques, it is likely that we missed other applications of AI for COVID-19. This review, therefore, might not identify all potential uses of AI for the current pandemic. Owing to practical constraints, the search was restricted to English studies. Therefore, we probably missed several studies written in other languages, especially Chinese. The search query did not include terms related to specific types of models or algorithms such as CNN, RNN, and SVM. Thus, it is likely that we missed some studies that used such terms in their title and abstract instead of the terms that we used (ie, AI, machine learning, and deep learning). The findings of this review are mostly based on preprints, which are more likely to have inaccurate or missing information. Therefore, the accuracy of the information in the included studies may affect the accuracy of our findings.

Conclusions

In this study, we provide a scoping review of 82 studies on AI against COVID-19. Given that many of the proposed methods are not yet clinically accepted, we remark that the most rewarding research will be on methods promising value beyond COVID-19. We believe that mobile phones offer unexploited potential, but more research in the direction of energy-efficient and federated learning is needed. We also believe that the use of NLP to assess effective communication of nonpharmaceutical interventions is a largely unexplored research direction, especially since data driving this research is available in the public domain, unlike much of the data produced by clinical studies. For AI to gain broad acceptance, standardized reporting protocols to be followed by studies on AI are needed. Likewise, more research on AI ethics and explainable AI is needed, paired with public education initiatives.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of artificial intelligence–based techniques.
[DOCX File , 14 KB - jmir_v22i12e20756_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 19 KB - jmir_v22i12e20756_app2.docx]

Multimedia Appendix 3
Interrater agreement matrices for study selection steps.
[DOCX File , 13 KB - jmir_v22i12e20756_app3.docx]

Multimedia Appendix 4
Data extraction form.
[DOCX File , 17 KB - jmir_v22i12e20756_app4.docx]

Multimedia Appendix 5
Characteristics of the included studies and features of artificial intelligence techniques used for COVID-19.
[DOCX File , 26 KB - jmir_v22i12e20756_app5.docx]

Multimedia Appendix 6
Features of data sets used for the development and validation of artificial intelligence models.
[DOCX File , 25 KB - jmir_v22i12e20756_app6.docx]

References

1. Kong W, Li Y, Peng M, Kong D, Yang X, Wang L, et al. SARS-CoV-2 detection in patients with influenza-like illness. Nat Microbiol 2020 May;5(5):675-678. [doi: 10.1038/s41564-020-0713-1] [Medline: 32265517]
2. Coronavirus disease (COVID-19): situation report– 121. World Health Organization. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200520-covid-19-sitrep-121.pdf [accessed 2020-04-19]
3. Chen Y, Li L. SARS-CoV-2: virus dynamics and host response. Lancet Infect Dis 2020 May;20(5):515-516 [FREE Full text] [doi: 10.1016/S1473-3099(20)30235-8] [Medline: 32213336]
4. Del Rio C, Malani PN. COVID-19-new insights on a rapidly changing epidemic. JAMA 2020 Apr 14;323(14):1339-1340. [doi: 10.1001/jama.2020.3072] [Medline: 32108857]
5. Fontanarosa PB, Bauchner H. COVID-19-looking beyond tomorrow for health care and society. JAMA 2020 May 19;323(19):1907-1908. [doi: 10.1001/jama.2020.6582] [Medline: 32301955]
6. Yang P, Wang X. COVID-19: a new challenge for human beings. Cell Mol Immunol 2020 May;17(5):555-557 [FREE Full text] [doi: 10.1038/s41423-020-0407-x] [Medline: 32235915]
7. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
8. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
9. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
10. van Hartskamp M, Consoli S, Verhaegh W, Petkovic M, van de Stolpe A. Artificial intelligence in clinical health care applications: viewpoint. Interact J Med Res 2019 Apr 05;8(2):e12100 [FREE Full text] [doi: 10.2196/12100] [Medline: 30950806]
11. Russell S, Norvig P. Artificial Intelligence: A Modern Approach, third edition. Upper Saddle River, NJ: Prentice Hall; 2010.
12. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. Physiol Genomics 2020 Apr 01;52(4):200-202 [FREE Full text] [doi: 10.1152/physiolgenomics.00029.2020] [Medline: 32216577]
13. Bullock J, Luccioni A, Hoffman Pham K, Sin Nga Lam C, Luengo-Oroz M. Mapping the landscape of artificial intelligence applications against COVID-19. arXiv. Preprint posted online March 25, 2020 . [doi: 10.1613/jair.1.12162]
14. Mondal B. Artificial intelligence: state of the art. In: Balas VE, Kumar R, Srivastava R, editors. Recent Trends and Advances in Artificial Intelligence and Internet of Things. New York, NY: Springer; 2019:389-425.
15. Fridman L. Deep learning state of the art (2020). YouTube. 2020. URL: https://www.youtube.com/watch?v=0VH1Lim8gL8&list=PLrAXtmErZgOeiKm4sgNOknGvNjby9efdf [accessed 2020-07-23]

16. ACMSIGGRAPH. Deep learning: a crash course. YouTube. 2018. URL: https://www.youtube.com/watch?v=r0Ogt-q956I [accessed 2020-07-23]
17. Santosh KC. AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data. J Med Syst 2020 Mar 18;44(5):93 [FREE Full text] [doi: 10.1007/s10916-020-01562-1] [Medline: 32189081]
18. Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. IEEE Rev Biomed Eng 2020 Apr 16;PP:1. [doi: 10.1109/RBME.2020.2987975] [Medline: 32305937]
19. Vaishya R, Javaid M, Khan I, Haleem A. Artificial intelligence (AI) applications for COVID-19 pandemic. Diabetes Metab Syndr 2020;14(4):337-339 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.012] [Medline: 32305024]
20. Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Systematic review and critical appraisal of prediction models for diagnosis and prognosis of COVID-19 infection. medRxiv. Preprint posted online April 5, 2020 . [doi: 10.1101/2020.03.24.20041020]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
22. Higgins J, Deeks J. Chapter 7: selecting studies and collecting data. In: Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, NJ: John Wiley & Sons; 2008.
23. Altman DG. Practical Statistics for Medical Research. Boca Raton, FL: CRC Press; 1990.
24. Barstugan M, Ozkaya U, Ozturk S. Coronavirus (COVID-19) classification using CT images by machine learning methods. arXiv. Preprint posted online March 20, 2020 .
25. Chen J, Wu L, Zhang J, Zhang L, Gong D, Zhao Y, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. medXriv. Preprint posted online March 1, 2020 . [doi: 10.1101/2020.02.25.20021568]
26. Fu M, Yi SL, Zeng Y, Ye F, Li Y, Dong X, et al. Deep learning-based recognizing COVID-19 and other common infectious diseases of the lung by chest CT scan images. medRxiv. Preprint posted online March 30, 2020 . [doi: 10.1101/2020.03.28.20046045]
27. Gozes O, Frid-Adar M, Greenspan H, Browning PD, Zhang H, Ji W, et al. arXiv. Preprint posted online March 10, 2020 .
28. Jin C, Chen W, Cao Y, Xu Z, Tan Z, Zhang X, et al. Development and evaluation of an AI system for COVID-19 diagnosis. medRxiv. Preprint posted online June 2, 2020 . [doi: 10.1101/2020.03.20.20039834]
29. Jin S, Wang B, Xu H, Luo C, Wei L, Zhao W, et al. AI-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system in four weeks. medRxiv. Preprint posted online March 23, 2020 . [doi: 10.1101/2020.03.19.20039354]
30. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology 2020 Aug;296(2):E65-E71 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
31. Shi F, Xia L, Shan F, Wu D, Wei Y, Yuan H, et al. Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification. arXiv. Preprint posted online March 22, 2020 .
32. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). medRxiv. Preprint posted online April 24, 2020 . [doi: 10.1101/2020.02.14.20023028]
33. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. medRxiv. Preprint posted online March 26, 2020 . [doi: 10.1101/2020.03.24.20042317]
34. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, et al. Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv. Preprint posted online February 21, 2020 .
35. Ying S, Zheng S, Li L, Zhang X, Zhang X, Huang Z, et al. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. medRxiv. Preprint posted online February 25, 2020 . [doi: 10.1101/2020.02.23.20026930]
36. Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. COVID-CT-Dataset: a CT scan dataset about COVID-19. arXiv. Preprint posted online March 30, 2020 .
37. Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. Deep learning-based detection for COVID-19 from chest CT using weak label. medRxiv. Preprint posted online March 26, 2020 . [doi: 10.1101/2020.03.12.20027185]
38. Zhou M, Chen Y, Yang D, Xu Y, Yao W, Huang J, et al. Improved deep learning model for differentiating novel coronavirus pneumonia and influenza pneumonia. medRxiv. Preprint posted online March 30, 2020 . [doi: 10.1101/2020.03.24.20043117]
39. Abbas A, Abdelsamea M, Gaber M. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. medRxiv. Preprint posted online May 18, 2020 . [doi: 10.1101/2020.03.30.20047456]
40. Apostolopoulos I, Mpesiana TA. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 2020 Jun;43(2):635-640 [FREE Full text] [doi: 10.1007/s13246-020-00865-4] [Medline: 32524445]

41. Bukhari SUK, Bukhari SSK, Syed A, Shah SSH. The diagnostic evaluation of convolutional neural network (CNN) for the assessment of chest x-ray of patients infected with COVID-19. medRxiv. Preprint posted online March 31, 2020 . [doi: 10.1101/2020.03.26.20044610]
42. Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, et al. Can AI help in screening Viral and COVID-19 pneumonia? arXiv. Preprint posted online March 29, 2020 .
43. Ghoshal B, Tucker A. Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection. arXiv. Preprint posted online March 22, 2020 .
44. Hemdan EED, Shouman MA, Karar ME. COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in x-ray images. arXiv. Preprint posted online March 24, 2020 .
45. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks. arXiv. Preprint posted online March 24, 2020 .
46. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with x-ray images. Comput Biol Med 2020 Jun;121:103792 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103792] [Medline: 32568675]
47. Sethy PK, Behera SK. Detection of coronavirus disease (COVID-19) based on deep features. Preprints. Preprint posted online March 19, 2020 . [doi: 10.20944/preprints202003.0300.v1]
48. Ucar F, Korkmaz D. COVIDiagnosis-Net: deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from x-ray images. Med Hypotheses 2020 Jul;140:109761 [FREE Full text] [doi: 10.1016/j.mehy.2020.109761] [Medline: 32344309]
49. Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. Sci Rep 2020 Nov 11;10(1):19549. [doi: 10.1038/s41598-020-76550-z] [Medline: 33177550]
50. Zhang J, Xie Y, Li Y, Shen C, Xia Y. COVID-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv. Preprint posted online March 27, 2020 .
51. Meng Z, Wang M, Song H, Guo S, Zhou Y, Li W, et al. Development and utilization of an intelligent application for aiding COVID-19 diagnosis. medRxiv. Preprint posted online March 21, 2020 . [doi: 10.1101/2020.03.18.20035816]
52. Feng C, Huang Z, Wang L, Chen X, Zhai Y, Zhu F, et al. A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected COVID-19 pneumonia in fever clinics. medRxiv. Preprint posted online March 20, 2020 . [doi: 10.1101/2020.03.19.20039099]
53. Lopez-Rincon A, Tonda A, Mendoza-Maldonado L, Claassen E, Garssen J, Kraneveld AD. Accurate identification of SARS-CoV-2 from viral genome sequences using deep learning. bioRxiv. Preprint posted online March 14, 2020 . [doi: 10.1101/2020.03.13.990242]
54. Wang Y, Hu M, Li Q, Zhang XP, Zhai G, Yao N. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner. arXiv. Preprint posted online February 12, 2020 .
55. Gao K, Nguyen DD, Wang R, Wei GW. Machine intelligence design of 2019-nCoV drugs. bioRxiv. Preprint posted online February 4, 2020 . [doi: 10.1101/2020.01.30.927889]
56. Ge Y, Tian T, Huang S, Wan F, Li J, Li S, et al. A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. bioRxiv. Preprint posted online March 12, 2020 . [doi: 10.1101/2020.03.11.986836]
57. Hofmarcher M, Mayr A, Rumetshofer E, Ruch P, Renz P, Schimunek J, et al. Large-scale ligand-based virtual screening for SARS-CoV-2 inhibitors using deep neural networks. SSRN J. Preprint posted online March 26, 2020 . [doi: 10.2139/ssrn.3561442]
58. Magar R, Yadav P, Farimani AB. Potential neutralizing antibodies discovered for novel corona virus using machine learning. arXiv. Preprint posted online March 18, 2020 . [doi: 10.1101/2020.03.14.992156]
59. Patankar S. Deep learning-based computational drug discovery to inhibit the RNA Dependent RNA Polymerase: application to SARS-CoV and COVID-19. OSF Preprints. Preprint posted online March 18, 2020 . [doi: 10.31219/osf.io/6kpbg]
60. Tang B, He F, Liu D, Fang M, Wu Z, Xu D. AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. bioRxiv. Preprint posted online March 8, 2020 . [doi: 10.1101/2020.03.03.972133]
61. Ton A, Gentile F, Hsing M, Ban F, Cherkasov A. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. Mol Inform 2020 Aug;39(8):e2000028 [FREE Full text] [doi: 10.1002/minf.202000028] [Medline: 32162456]
62. Zhang H, Saravanan KM, Yang Y, Hossain MD, Li J, Ren X, et al. Deep learning based drug screening for novel coronavirus 2019-nCov. Preprints. Preprint posted online February 5, 2020 . [doi: 10.20944/preprints202002.0061.v1]
63. Zhavoronkov A, Aladinskiy V, Zhebrak A, Zagribelnyy B, Terentiev V, Bezrukov DS, et al. Potential 2019-nCoV 3C-like protease inhibitors designed using generative deep learning approaches. ChemRxiv. Preprint posted online February 10, 2020 . [doi: 10.26434/chemrxiv.11829102]
64. Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model. bioRxiv. Preprint posted online February 2, 2020 . [doi: 10.1101/2020.01.31.929547]

65. Hu F, Jiang J, Yin P. Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model. arXiv. Preprint posted online March 2, 2020 .
66. Wang Z, Li L, Yan J, Yao Y. Evaluating the traditional Chinese medicine (TCM) officially recommended in China for COVID-19 using ontology-based side-effect prediction framework (OSPF) and deep learning. Preprints. Preprint posted online February 17, 2020 . [doi: 10.20944/preprints202002.0230.v1]
67. Abdelmageed MI, Abdelmoneim AH, Mustafa MI, Elfadol NM, Murshed NS, Shantier SW, et al. Design of multi epitope-based peptide vaccine against E protein of human 2019-nCoV: an immunoinformatics approach. bioRxiv. Preprint posted online February 11, 2020 . [doi: 10.1101/2020.02.04.934232]
68. Ong E, Wong MU, Huffman A, He Y. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. bioRxiv. Preprint posted online March 21, 2020 . [doi: 10.1101/2020.03.20.000141]
69. Rahman MS, Hoque MN, Islam MR, Akter S, Rubayet-Ul-Alam ASM, Siddique MA, et al. Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2 etiologic agent of global pandemic COVID-19: an in silico approach. bioRxiv. Preprint posted online March 31, 2020 . [doi: 10.1101/2020.03.30.015164]
70. Sarkar B, Ullah MA, Johora FT, Taniya MA, Araf Y. The essential facts of Wuhan novel coronavirus outbreak in China and epitope-based vaccine designing against 2019-nCoV. bioRxiv. Preprint posted online February 11, 2020 . [doi: 10.1101/2020.02.05.935072]
71. Fast E, Altman RB, Chen B. Potential T-cell and B-cell epitopes of 2019-nCoV. bioRxiv. Preprint posted online February 21, 2020 . [doi: 10.1101/2020.02.19.955484]
72. Heo L, Feig M. Modeling of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins by machine learning and physics-based refinement. bioRxiv. Preprint posted online March 28, 2020 . [doi: 10.1101/2020.03.25.008904]
73. Qiao R, Tran NH, Shan B, Ghodsi A, Li M. Personalized workflow to identify optimal T-cell epitopes for peptide-based vaccines against COVID-19. arXiv. Preprint posted online March 24, 2020 .
74. Demirci MDS, Adan A. Computational analysis of microRNA-mediated interactions in SARS-CoV-2 infection. bioRxiv. Preprint posted online March 17, 2020 . [doi: 10.1101/2020.03.15.992438]
75. Al-Qaness MAA, Ewees AA, Fan H, Abd El Aziz M. Optimization method for forecasting confirmed cases of COVID-19 in China. J Clin Med 2020 Mar 02;9(3):674 [FREE Full text] [doi: 10.3390/jcm9030674] [Medline: 32131537]
76. Dandekar R, Barbastathis G. Neural network aided quarantine control model estimation of COVID spread in Wuhan, China. arXiv. Preprint posted online March 18, 2020 .
77. Dutta S, Bandyopadhyay SK. Machine learning approach for confirmation of COVID-19 cases: positive, negative, death and release. medRxiv. Preprint posted online March 30, 2020 . [doi: 10.1101/2020.03.25.20043505]
78. Fong SJ, Li G, Dey N, Crespo RG, Herrera-Viedma E. Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. arXiv. Preprint posted online March 22, 2020 .
79. Fong SJ, Li G, Dey N, Crespo RG, Herrera-Viedma E. Finding an accurate early forecasting model from small dataset: a case of 2019-nCoV novel coronavirus outbreak. arXiv. Preprint posted online March 24, 2020 .
80. Hu Z, Ge Q, Li S, Boerwincle E, Jin L, Xiong M. Forecasting and evaluating intervention of covid-19 in the world. arXiv. Preprint posted online March 22, 2020 .
81. Hu Z, Ge Q, Li S, Jin L, Xiong M. Artificial intelligence forecasting of covid-19 in China. arXiv. Preprint posted online February 17, 2020 .
82. Huang CJ, Chen YH, Ma Y, Kuo PH. Multiple-input deep convolutional neural network model for COVID-19 forecasting in China. medRxiv. Preprint posted online March 27, 2020 . [doi: 10.1101/2020.03.23.20041608]
83. Kumar P, Kalita H, Patairiya S, Sharma YD, Nanda C, Rani M, et al. Forecasting the dynamics of COVID-19 pandemic in top 15 countries in April 2020: ARIMA model with machine learning approach. medRxiv. Preprint posted online April 5, 2020 . [doi: 10.1101/2020.03.30.20046227]
84. Li M, Zhang Z, Jiang S, Liu Q, Chen C, Zhang Y, et al. Predicting the epidemic trend of COVID-19 in China and across the world using the machine learning approach. medRxiv. Preprint posted online April 20, 2020 . [doi: 10.1101/2020.03.18.20038117]
85. Marini M, Chokani N, Abhari RS. COVID-19 epidemic in Switzerland: growth prediction and containment strategy using artificial intelligence and big data. medRxiv. Preprint posted online April 7, 2020 . [doi: 10.1101/2020.03.30.20047472]
86. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill 2020 Mar;25(10):2000180 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000180] [Medline: 32183930]
87. Pirouz B, Shaffiee Haghshenas S, Shaffiee Haghshenas S, Piro P. Investigating a serious challenge in the sustainable development process: analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. Sustainability 2020 Mar 20;12(6):2427. [doi: 10.3390/su12062427]
88. Tiwari S, Kumar S, Guleria K. Outbreak trends of coronavirus disease-2019 in India: a prediction. Disaster Med Public Health Prep 2020 Apr 22:1-6 [FREE Full text] [doi: 10.1017/dmp.2020.115] [Medline: 32317044]
89. Guo Q, Li M, Wang C, Wang P, Fang Z, Tan J, et al. Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm. bioRxiv. Preprint posted online January 24, 2020 .

90. Qiang X, Xu P, Fang G, Liu W, Kou Z. Using the spike protein feature to predict infection risk and monitor the evolutionary dynamic of coronavirus. Infect Dis Poverty 2020 Mar 25;9(1):33 [FREE Full text] [doi: 10.1186/s40249-020-00649-8] [Medline: 32209118]
91. Randhawa GS, Soltysiak MPM, El Roz H, de Souza CPE, Hill KA, Kari L. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. bioRxiv. Preprint posted online February 4, 2020 . [doi: 10.1101/2020.02.03.932350]
92. Chen X, Yao L, Zhang Y. Residual attention U-Net for automated multi-class segmentation of COVID-19 chest CT images. arXiv. Preprint posted online April 12, 2020 .
93. Gaál G, Maga B, Lukács A. Attention U-Net based adversarial architectures for chest x-ray lung segmentation. arXiv. Preprint posted online March 23, 2020 .
94. Huang L, Han R, Ai T, Yu P, Kang H, Tao Q, et al. Serial quantitative chest CT assessment of COVID-19: a deep learning approach. Radiol Cardiothoracic Imaging 2020 Apr 01;2(2):e200075. [doi: 10.1148/ryct.2020200075]
95. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Lung infection quantification of COVID-19 in CT images with deep learning. arXiv. Preprint posted online March 10, 2020 .
96. Tang Z, Zhao W, Xie X, Zhong Z, Shi F, Liu J, et al. Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. arXiv. Preprint posted online March 26, 2020 .
97. Yu H, Shao J, Guo Y, Xiang Y, Sun C, Yuan Y. Data-driven discovery of a clinical route for severity detection of COVID-19 pediatric cases. medRxiv. Preprint posted online March 20, 2020 . [doi: 10.1101/2020.03.09.20032219]
98. Bai X, Fang C, Zhou Y, Bai S, Liu Z, Chen Q, et al. Predicting COVID-19 malignant progression with AI techniques. medRxiv. Preprint posted online March 23, 2020 . [doi: 10.1101/2020.03.20.20037325]
99. DeCaprio D, Gartner J, McCall CJ, Burgess T, Kothari S, Sayed S. Building a COVID-19 vulnerability index. medRxiv. Preprint posted online March 21, 2020 . [doi: 10.1101/2020.03.16.20036723]
100. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A tool to early predict severe 2019-novel coronavirus pneumonia (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. medRxiv. Preprint posted online March 20, 2020 . [doi: 10.1101/2020.03.17.20037515]
101. Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using artificial intelligence to help medical decision-making. medRxiv. Preprint posted online April 1, 2020 . [doi: 10.1101/2020.03.30.20047308]
102. Yan L, Zhang HT, Goncalves J, Xiao Y, Wang M, Guo Y, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. medRxiv. Preprint posted online March 17, 2020 . [doi: 10.1101/2020.02.27.20028027]
103. Sarkar J, Chakrabarti P. A machine learning model reveals older age and delayed hospitalization as predictors of mortality in patients with COVID-19. medRxiv. Preprint posted online March 30, 2020 . [doi: 10.1101/2020.03.25.20043331]
104. Qi X, Jiang Z, Yu Q, Shao C, Zhang H, Yue H, et al. Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. medRxiv. Preprint posted online March 3, 2020 . [doi: 10.1101/2020.02.29.20029603]
105. Pandey R, Gautam V, Bhagat K, Sethi T. A machine learning application for raising WASH awareness in the times of covid-19 pandemic. arXiv. Preprint posted online March 16, 2020 .
106. Rolling updates on coronavirus disease (COVID-19). World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2020-05-21]
107. Konečný J, McMahan B, Ramage D. Federated optimization: distributed optimization beyond the datacenter. arXiv. Preprint posted online November 11, 2015 .
108. Kiwibot. 2020. URL: https://www.kiwibot.com/ [accessed 2020-05-22]
109. Whiteboard Coordinator. 2020. URL: https://wbcoordinator.com/ [accessed 2020-05-22]
110. EHTERAZ. App Store. 2020. URL: https://apps.apple.com/us/app/ehteraz/id1507150431 [accessed 2020-05-22]

Abbreviations

AI: artificial intelligence
CNN: convolutional neural network
CT: computed tomography
NCBI: National Center for Biotechnology Information
NLP: natural language processing
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RNN: recurrent neural network
SVM: support vector machine

Edited by G Eysenbach; submitted 27.05.20; peer-reviewed by N Fareed, KL Ong; comments to author 13.07.20; revised version received 26.07.20; accepted 29.07.20; published 15.12.20.

Please cite as:
Abd-Alrazaq A, Alajlani M, Alhuwail D, Schneider J, Al-Kuwari S, Shah Z, Hamdi M, Househ M
Artificial Intelligence in the Fight Against COVID-19: Scoping Review
J Med Internet Res 2020;22(12):e20756
URL: http://www.jmir.org/2020/12/e20756/
doi:10.2196/20756
PMID:33284779

Original Paper

Limitations of Deep Learning Attention Mechanisms in Clinical Research: Empirical Case Study Based on the Korean Diabetic Disease Setting

Junetae Kim[1,2,3*], PhD; Sangwon Lee[2*], MS; Eugene Hwang[4], BS; Kwang Sun Ryu[2], PhD; Hanseok Jeong[1], MS; Jae Wook Lee[5], MD, PhD; Yul Hwangbo[3,6], MD; Kui Son Choi[1,2], PhD; Hyo Soung Cha[1,2], PhD

[1]Graduate School of Cancer Science and Policy, National Cancer Center, Goyang-si, Gyeonggi-do, Republic of Korea
[2]Cancer Data Center, National Cancer Control Institute, National Cancer Center, Goyang-si, Gyeonggi-do, Republic of Korea
[3]Healthcare AI Team, Healthcare Platform Center, National Cancer Center, Goyang-si, Gyeonggi-do, Republic of Korea
[4]School of Management Engineering, Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea
[5]Division of Nephrology, Department of Internal Medicine, National Cancer Center, Goyang-si, Gyeonggi-do, Republic of Korea
[6]Division of Endocrinology, Department of Internal Medicine, National Cancer Center, Goyang-si, Gyeonggi-do, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Hyo Soung Cha, PhD
Cancer Data Center
National Cancer Control Institute
National Cancer Center
809 Madu 1(il)-dong, Ilsandong-gu
Goyang-si, Gyeonggi-do, 10408
Republic of Korea
Phone: 82 31 920 1892
Email: kkido@ncc.re.kr

Abstract

Background: Despite excellent prediction performance, noninterpretability has undermined the value of applying deep-learning algorithms in clinical practice. To overcome this limitation, attention mechanism has been introduced to clinical research as an explanatory modeling method. However, potential limitations of using this attractive method have not been clarified to clinical researchers. Furthermore, there has been a lack of introductory information explaining attention mechanisms to clinical researchers.

Objective: The aim of this study was to introduce the basic concepts and design approaches of attention mechanisms. In addition, we aimed to empirically assess the potential limitations of current attention mechanisms in terms of prediction and interpretability performance.

Methods: First, the basic concepts and several key considerations regarding attention mechanisms were identified. Second, four approaches to attention mechanisms were suggested according to a two-dimensional framework based on the degrees of freedom and uncertainty awareness. Third, the prediction performance, probability reliability, concentration of variable importance, consistency of attention results, and generalizability of attention results to conventional statistics were assessed in the diabetic classification modeling setting. Fourth, the potential limitations of attention mechanisms were considered.

Results: Prediction performance was very high for all models. Probability reliability was high in models with uncertainty awareness. Variable importance was concentrated in several variables when uncertainty awareness was not considered. The consistency of attention results was high when uncertainty awareness was considered. The generalizability of attention results to conventional statistics was poor regardless of the modeling approach.

Conclusions: The attention mechanism is an attractive technique with potential to be very promising in the future. However, it may not yet be desirable to rely on this method to assess variable importance in clinical settings. Therefore, along with theoretical studies enhancing attention mechanisms, more empirical studies investigating potential limitations should be encouraged.

(J Med Internet Res 2020;22(12):e18418) doi:10.2196/18418

KEYWORDS

attention; deep learning; explainable artificial intelligence; uncertainty awareness; Bayesian deep learning; artificial intelligence; health data

Introduction

In recent years, there has been significant evidence that deep-learning algorithms can outperform other machine-learning algorithms and conventional statistics in the medical field [1,2]. Despite the better prediction accuracy than conventional algorithms, the implications of using deep learning have been limited owing to the inability to explain the models [3,4]. Particularly in the medical environment, where the association between a disease and symptoms must be identified to provide adequate treatments, the interpretability of models is very important [3-5]. To overcome these limitations, interpretable deep-learning algorithms such as Shapley Additive Explanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and attention mechanisms have been introduced [6-8]. The commonality of these three methodologies provides interpretability in the form of variable importance [6-8]. The difference between the methodologies is that with SHAP or LIME, variable importance is measured through simulations that change the data after model training is completed [6,7], whereas under attention mechanisms, variable importance is inferred during model training, which improves model performance by weighting several important variables [8,9].

Based on this advantage, attention mechanisms have starting to gain appeal in the clinical research field [10-17]. However, there is a gap between the application of attention mechanisms in clinical research and up-to-date attention algorithms in development. Specifically, most of the recent attention studies have focused on improving the theoretical robustness, design approach, and model accuracy with attention mechanisms [10-17]. However, clinical researchers are more interested in potential limitations that may arise when attention mechanisms are applied, and how they may differ from conventional statistics, than in the details as to how robust and sophisticated attention mechanisms are being developed. A few studies have introduced the potential limitations of attention mechanisms [18,19]. However, these studies have been theoretical, making it difficult for clinical researchers to understand and accept the results. Thus, it is increasingly necessary to provide a discussion of what clinical researchers should be aware of when applying the new concept of attention mechanisms in their research.

With the goal of reducing this gap, the aim of this study was to evaluate attention mechanisms in terms of prediction performance and interpretability. In addition, there remains a lack of guidance for clinical researchers in the implementation of attention mechanisms; therefore, to facilitate understanding for clinical researchers, this study preemptively provides basic concepts, key considerations, and codes for attention mechanisms. Finally, a case analysis was performed in a cross-sectional and structured data environment, which is the simplest data setting possible for clinical researchers.

This study was conducted according to the following procedure. First, the scope of the study was established in terms of the data structure. Then, a brief introduction and several important considerations regarding attention mechanisms were considered. Second, based on previous research, a two-dimensional framework was established to guide the four modeling approaches to attention mechanisms. Third, five empirical tests with attention mechanisms were performed using the four models: prediction performance, probability reliability, concentration of variable importance, consistency of attention results, and generalizability of attention results to conventional statistics. Finally, potential limitations that may arise when using attention mechanisms were identified.

Methods

Research Scope

Since the design approaches of attention mechanisms differ greatly depending on the data structure, the scope of this study was established in terms of data structure. Specifically, attention mechanism research in the medical field can be divided into two main categories from a data point of view. The first category is an unstructured data area where data containing natural language and images cannot be stored in a row and column table structure [20,21]. In the field of natural language processing, attention mechanisms have been applied to determine the relationship between words or between words and diseases in clinical notes [13,14,22]. In the image area, attention mechanisms have been used to highlight which parts of clinical images were related to clinical events, or to annotate the images [15-17]. The second category is a structured data area where data can be organized in table formats with a row and column structure [21]. In this area, attention mechanisms have been applied to electronic health records to determine variables that are strongly associated with clinical events [9,12,23].

Structured data familiar to clinical researchers are widely applicable to most statistical analyses, including linear regression analysis and analysis of variance (ANOVA). Since one purpose of this study was to compare the results of attention mechanisms and conventional statistical methods, the scope of the study was limited to structured data. Furthermore, most previous attention mechanism studies using structured data have been conducted in time-series settings [9,12,23]. However, this study was conducted in a cross-sectional data setting, which is simpler and easier than a time-series data setting, and can therefore help readers less familiar with attention mechanisms to better understand the results of the case study.

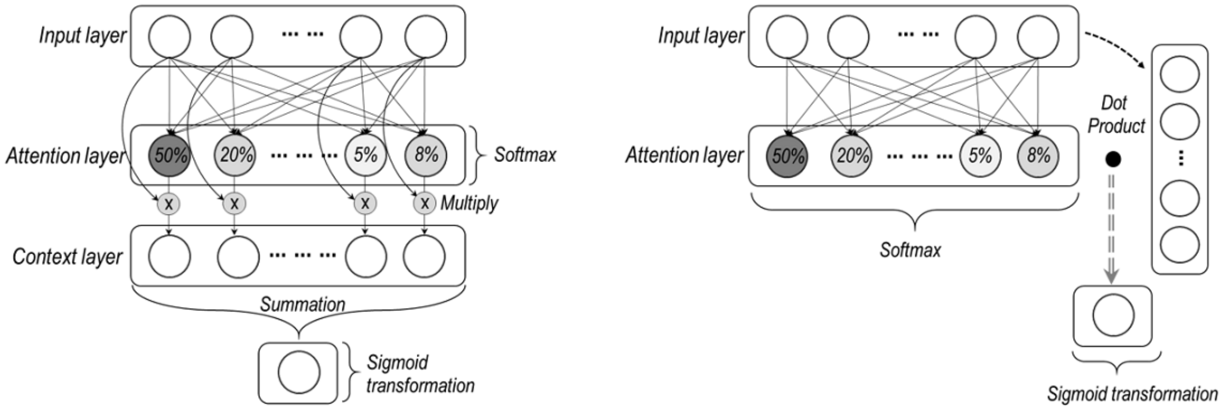
Introduction to Attention

Concepts of Attention Mechanisms

Attention, one of the layers in a neural network model, quantifies the importance of input variables in terms of their impact on outcomes (Figure 1) [8,24,25]. Attention is mostly calculated based on the Softmax function (Notation A1 in Multimedia Appendix 1), such that each node in the layer has a value between 0 and 1 and the sum of all node values must be 1

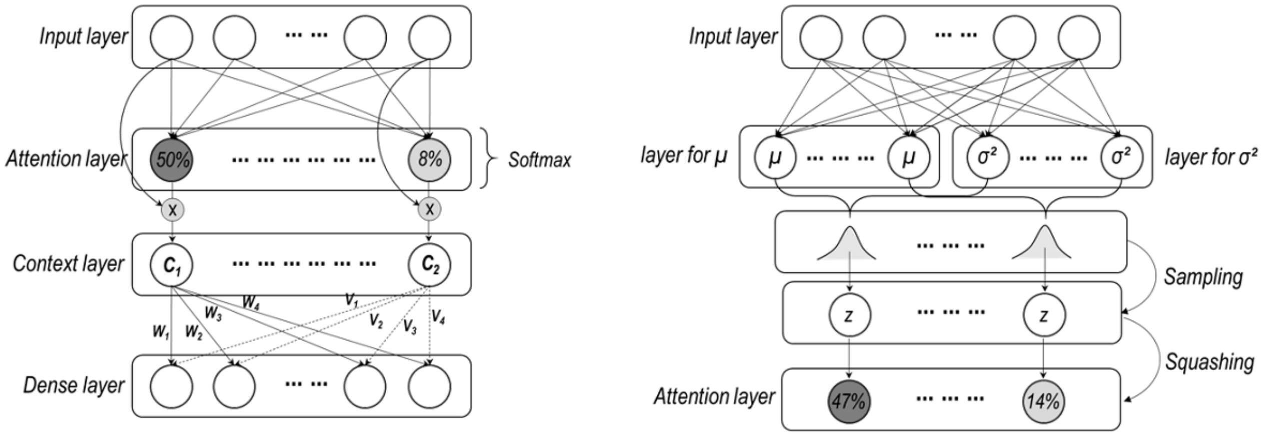
[8,24,25]. When the node size of attention is equal to the number of input variables, the influence of the input variables can be transferred toward the model outcome by multiplying the attention values with the corresponding input variables (Context layer in Figure 1A) [8,24,25]. Accordingly, in the case of binary classification, all values in the context layer are summed together to produce a single value (Summation in Figure 1A). The efficient model design without context layers is possible through the dot product between the input and attention values (Figure 1B). Finally, the single value may be converted to a value between 0 and 1 through the sigmoid function (Sigmoid transformation in Figure 1A and B).

Figure 1. Model structures for attention implementation. (A) Basic architecture of an attention mechanism model. (B) Model architecture where the dot product is employed for transferring the influence of input variables toward the outcome. (C) Model architecture where the importance of input variables may be decayed. (D) Model architecture that is aware of uncertainty. The percentages in the circles show examples of attention values.



(A) Basic architecture

(B) Dot product architecture



(C) Importance decayed architecture

(D) Uncertainty awareness architecture

The attention value of a certain variable indicates the relative importance of that variable compared with that of other variables. If the attention value of a particular variable is large, the large influence of that variable is transmitted when predicting the outcome variable. As an extreme example, when the attention value of a variable is 1, only that variable is used to predict the outcome variable, whereas if the attention value of a variable is 0, that variable is not used to predict the outcome variable. Figure 1A shows the basic architecture of a model with attention mechanisms; the code for the model implemented in Keras is provided in Codes A1 of Multimedia Appendix 2.

Consideration in Attention Modeling

Attention mechanisms can be implemented in various ways, because the key feature of deep-learning modeling is that users can freely design the structure [8,9,23-25]. However, there is a primary important consideration in implementing attention mechanisms. In some cases, the influence of variable importance in the context layer can be distorted. For instance, if all w_i values are close to 0, the value of C_1 has a minor effect on the next layer even if that value is the highest in the context layer (Figure 1C). Moreover, even if the value of C_2 is the lowest, if all v_is have very large positive values, the large influence of C_2 can be passed to the next layer (Figure 1C). As such, context values can be skewed as they are computed through a weight matrix in the process of being passed to the next layer (ie, Dense layer in Figure 1C). As a result, the skewed effects can be propagated to the model output if the output is inferred from the layer. Therefore, it is very important to design a structure where the outputs are not computed through weight matrices [9,23].

Modeling Approaches

Although deep-learning models can be developed in various ways depending on the tendency of developers, two approaches have been commonly applied in recent attention studies: increasing the degrees of freedom and uncertainty awareness (UA).

Increase in the Degrees of Freedom

The mechanism for increasing the degrees of freedom is to design multi-attention layers; representative algorithms that reflect such a mechanism include transformer and bidirectional transformer (BERT) [8,26]. Our intuition regarding the effectiveness of the mechanism relies on the idea that models can learn the importance of input variables from various perspectives [8,26]. Given the randomness feature of deep-learning training, the result from one attention layer can be unreliable. However, the multi-attention model offers multiple result sets with variable importance so that a reliable set of results may compensate for an unreliable set. Consequently, models in which multi-attention layers are applied have recently shown better performance than other models [8,26-28].

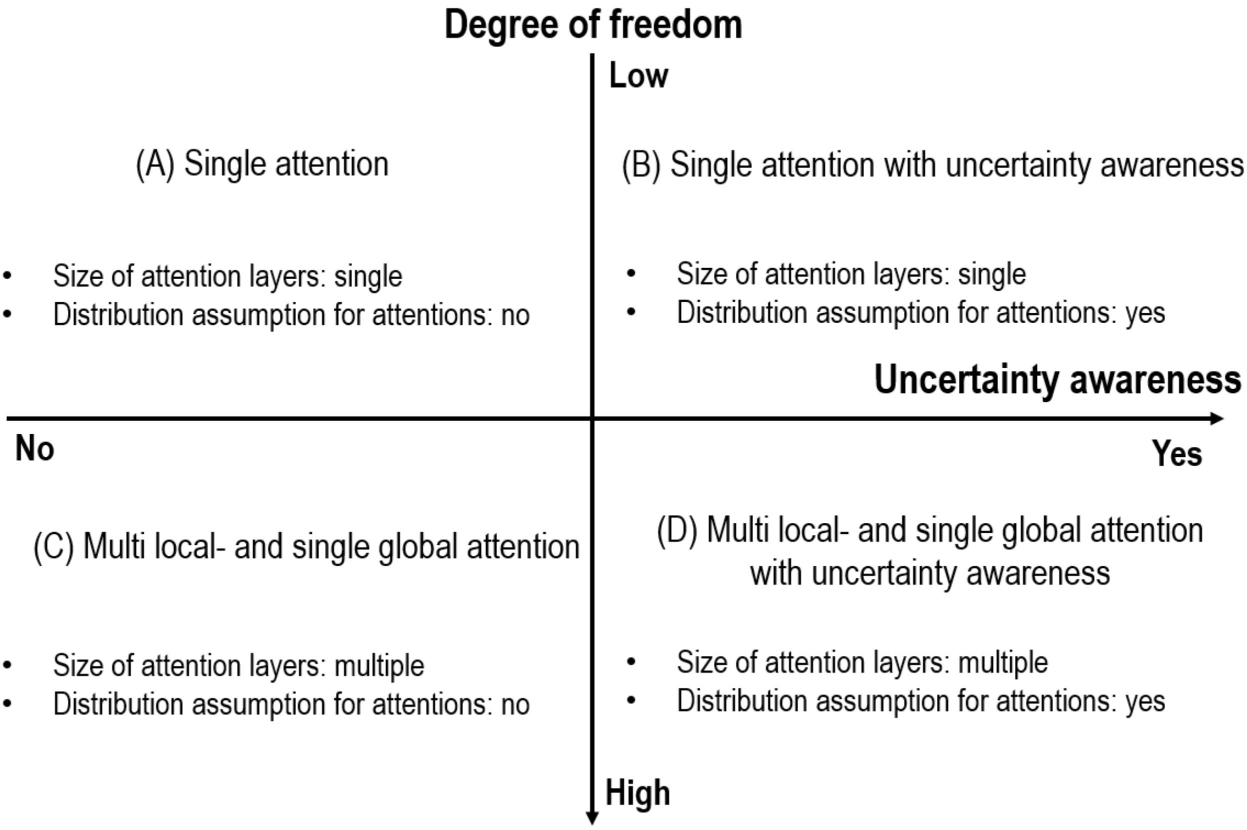
UA

Deep-learning algorithms are not free from the uncertainty issue, which concerns the fact that prediction results have the potential to be incomplete in terms of accuracy and consistency [9,29-33]. The major sources of uncertainty include data with noise and omissions, the complexity of the model associated with the parameters (ie, number of weights and type of activation functions), and the structures (ie, degree of depth) [30,31]. One way to alleviate this issue is to consider the presence of uncertainty in modeling [9,29,32,33]. Specifically, we may assume that node values (ie, attention values) in a certain layer come from a distribution with a mean (μ) and a variance (σ^2; Figure 1D) [9,29,32,33]. A normal distribution (ie, a Gaussian distribution) that is theoretically clear and can be computed efficiently is often assumed [9,29,32,33]. Based on this assumption, certain values with high probability are estimated, which may mitigate the random nature of deep-learning training [9,29-33]. A representative model designed under these assumptions is the variational auto-encoder [32,33].

Framework for Empirical Tests

Based on the discussion above, two directions (ie, degree of freedom and UA) were considered for attention modeling (Figure 2). In this two-dimensional framework, four cases were suggested for empirical tests (Figure 2). Degree of freedom is related to model structures and UA is related to the estimation approach.

Figure 2. Framework for empirical tests.



Empirical test entries for the four models in the framework were categorized into two broad categories: outcome and attention (Table 1). In terms of model outcome, a receiver operating characteristic (ROC) test, which expresses model accuracy based on the relationship between sensitivity and specificity, was employed for prediction performance [34]. In addition, the performance of probability reliability, which measures the degree of agreement between predicted and actual probability, was assessed using a reliability diagram and Brier scores [35-37].

In terms of attention, the degree of how concentrated the variable importance was in particular variables was measured (ie, Concentration in Table 1). The Herfindahl index, which represents the degree of concentration with values ranging from near 0 (least concentrated) to 1 (most concentrated), was employed for this measure [38]. Furthermore, correlation analysis was conducted to evaluate the consistency of attention results between multiple instances. Lastly, the generalizability of attention results was tested in two ways. First, the variable effect sizes obtained from conventional statistical methods (*t* test, Cohen *d*; chi-square test, Cramer *V*) were compared with variable importance (ie, attention values) [39,40]. For clear comparison from a clinical point of view, only the top 5% of the variables in terms of effect size (ie, conventional methods) and variable importance (ie, attention) were compared. Second, regression analysis was used to determine the overall relationship between attention values and effect sizes from conventional methods.

Table 1. Empirical test entries for measuring the performance of four models.

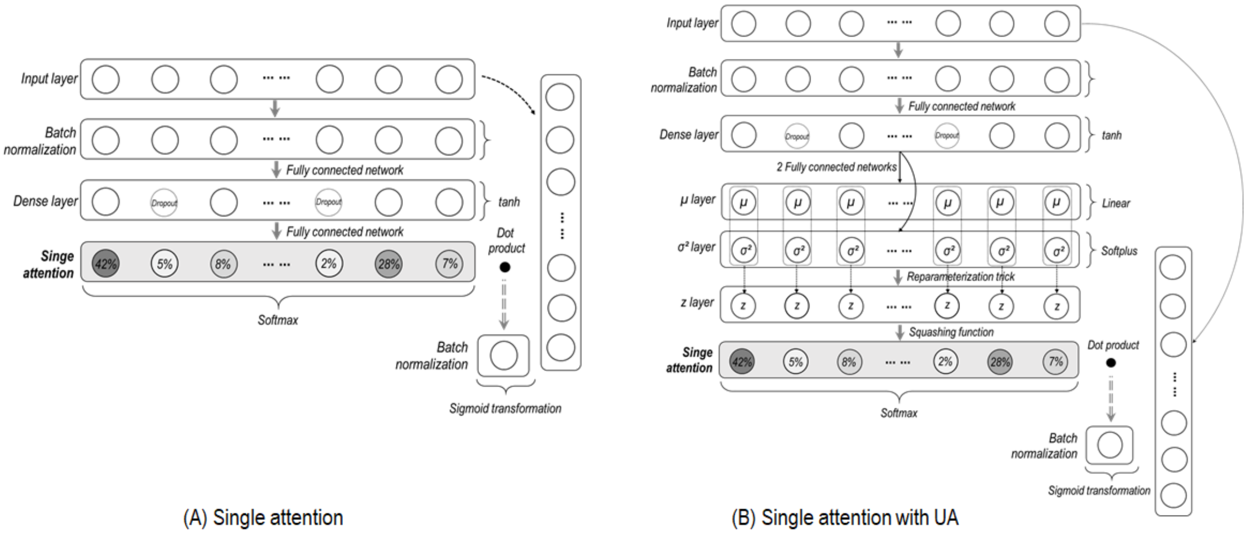
Entries (measures)	Methods
Outcome	
Prediction performance	- Receiver operating characteristic
Probability reliability	- Reliability diagrams
Attention	
Concentration	- Herfindahl index (near 0, least concentrated; 1, most concentrated)
Consistency	- Correlation
Generalizability	- Effect size: Cohen *d* (*t* test), Cramer *V* (chi-square test) - Regression analysis (dependent variable, effect size obtained from conventional methods; independent variable, attention values)

Model Specifications

Model Designs

Four models were developed according to the framework presented in Figure 2. The letters A, B, C, and D represent quadrants on the framework that correspond to the letters representing the model designs in Figure 3 and Figure 4. Model A (without any uncertainty considerations) has only a single attention layer (Figure 3A). The basic design of model B is the same as that of model A; however, it differs in that it has additional layers for UA (see layers with μ, σ^2, and z in Figure 3B). Thus, attention values in model B were estimated from the Gaussian distribution [9,33].

Figure 3. Model designs for single attention mechanisms. UA: uncertainty awareness. The concept of "reparameterization trick" is described in Concept A1 of Multimedia Appendix 1.



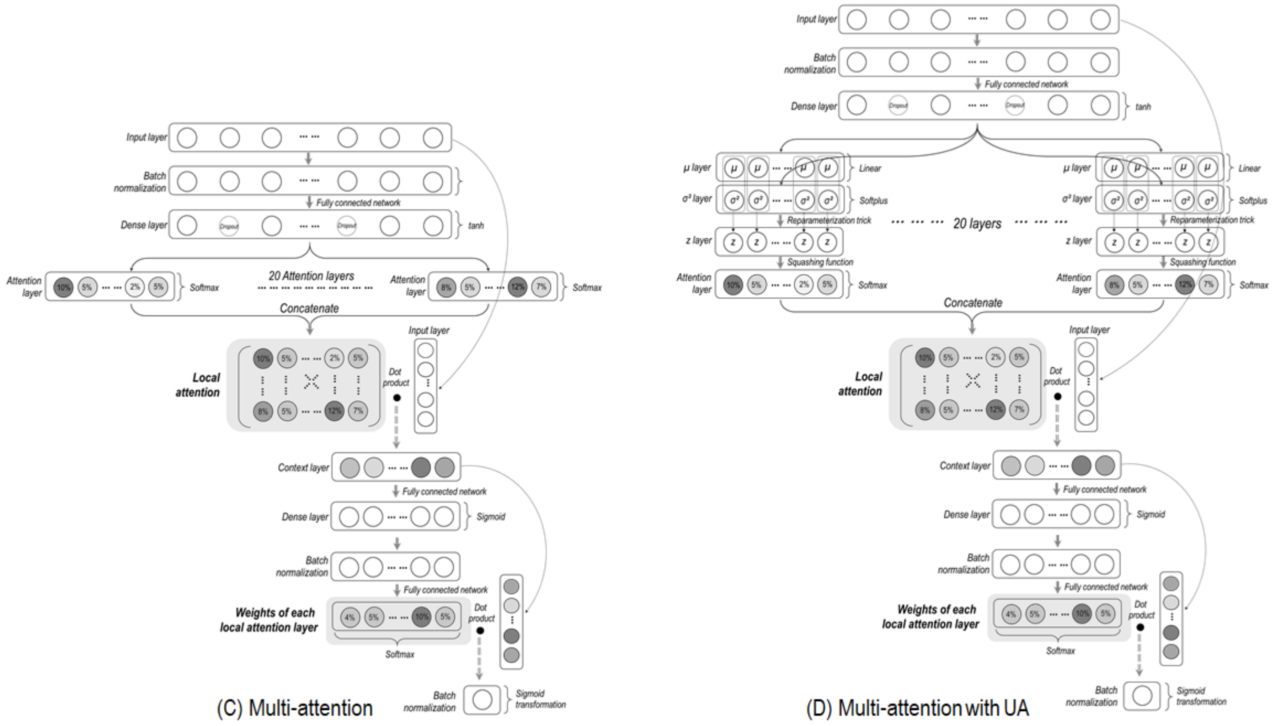
(A) Single attention

(B) Single attention with UA

The two models in which the degree of freedom is considered are presented in Figure 4. The difference between models C and D is that uncertainty in attention estimation is considered in model D (see layers with μ, σ^2, and z in Figure 4D).

Since these models have multi-attention layers (ie, Local attention in Figure 4) with a heuristic size of 20, multiple attentions are estimated. Thus, a novel structure was designed to convey the multiple values in the direction of model outcomes. Specifically, a context layer was created as the dot product of the local attention layer and the input layer (see Context layer in Figure 4). Each value on the context layer represents the summed impact of each corresponding local attention layer. Next, a “Weights of each local attention layer” was formed, whose role is to weigh (with weights between 0 and 1) the summed impact values in the context layer (Figure 4). Lastly, the outcome layer was created as the dot product of the weights of each local attention layer and the context layer.

Figure 4. Model designs for multi-attention mechanisms. UA: uncertainty awareness. The concept of "reparameterization trick" is described in Concept A1 of Multimedia Appendix 1.



(C) Multi-attention

(D) Multi-attention with UA

This somewhat complex structure ensures that the influence of one variable is passed only once to the model outcome, even if attention values are inferred multiple times (20 times in this case). Furthermore, using both the local attention layer and the weights corresponding to each vector, a unique attention value for each variable can be obtained, which facilitates interpretation.

Graphical and mathematical notations are provided for obtaining a set of unique values (global attention in Figure A1 of Multimedia Appendix 1). In addition, details of the four models are provided as Keras codes in Codes A2-A5 of Multimedia Appendix 2.

Settings for Rigorous Analysis

A 10-fold test was performed to assess the empirical test entries. The dataset was divided into 10 test sets (10% of total sets) and 10 training sets (90% of total sets). The training sets were then subdivided with 80% used directly for model training and 20% for validation.

Entries related to the model outcome (Table 1) were evaluated using all predicted probabilities of the entire sample. In other words, all of the values estimated from the 10 test sets were combined into one total set, which was then used for testing. Entries related to model attention were assessed based on the level of fold sets. Specifically, all of the estimated attention values were aggregated so that each set of 10 folds had a representative value. For rigorous estimation, both outcome and attention were estimated through Monte Carlo simulations with 100 trials [5,25]. Detailed procedures for estimating outcome and attention values are provided in the form of pseudoalgorithms (Algorithms A1-A3) in Multimedia Appendix 2.

Cost Functions

The binary cross-entropy function, which is generally used in binary outcome settings [41,42], was employed for models A and C where UA was not considered. However, the loss functions for models B and D, given their UA, were specified differently.

The UA models assume that the model outcome is dependent on the normal distribution (ie, layers with μ, σ^2, and z), which infers the attention value [9,32,33]. Therefore, the distribution associated with attention should be considered in the cost function. The cost function under these assumptions was derived through Bayesian inference theory [9,32,33,43]. According to this theory, the network weights should be learned so that the distributions in the *z* layer generated by the weights (see *z* layer in Figure 3B) become similar to the true distributions in the *z* layer [9,32,33,43]. Therefore, the cost function for uncertainty awareness models consists of two terms: the loss associated with the model outcome and the degree of similarity associated with the *z* distribution [9,32,33,43]. The cost function, with its description, is presented in Notation A2 of Multimedia Appendix 1.

Learning Environments and Parameters

Attention models were developed, learned, and tested on Keras 2.3.1, tensorflow 2.1.0, and Python 3.7.6. Adam with a learning rate of 0.001 was employed as an optimizer to train all models. A training dataset with a batch size of 5000 was provided to the model. The early-stop rule was applied to stop training the models at the optimal epoch. Thus, model training was terminated when the loss value of the validation set did not improve further during the last 1200 epochs. Other details about activation functions and the structure of nodes and layers are provided in Code A2 to Code A5 of Multimedia Appendix 2. For the effect sizes of conventional statistical methods, the values for Cohen *d* and Cramer *V* were obtained from researchpy 0.2.3, a third-party Python library. Additionally, regression analysis was performed on Stata 13, a commercial statistical analysis software.

Data

The case analysis was performed in a setting where the relationship between a disease and other variables is well established: an 8-year (2010 to 2017 inclusive) cumulative Korea National Health and Nutrition Examination Survey dataset, which assesses the nutrition and health status of Koreans and collects information about major chronic diseases such as metabolic syndrome and diabetes [44]. Since the association between diabetes and other variables has been well established through prior studies using these data, this selection facilitated a clear assessment of the empirical test results of this study [45-49]. Thus, a diabetes diagnosis (1=diabetes, 0=no diabetes) was set as the outcome variable for the four attention models. The subjects were classified as having or not having diabetes based on whether they were diagnosed by a doctor, or received diabetes medication or insulin injections. Fasting blood glucose levels, which are a very strong indicator for diagnosing diabetes, were intentionally used as an input variable to evaluate the power of the attention mechanism for determining important variables.

In the 8-year cumulative data, only variables with consistent labels during that period were included. Variables with no change in value, containing more than a 50% omission rate and subject identification information were excluded from the study set. Categorical variables of both nominal and ordinal types were integerized using integer encoding [50]. In other words, class labels of each categorical variable were converted into integers. Missing values were encoded as the extreme value 99,999. Since deep-learning algorithms can learn the nonlinear relations among variables [51], these encoding approaches can be workable and are efficient in settings where preprocessing is demanding owing to many variables. All values in input variables except for a missing value indicator (ie, 99,999) were normalized to be between 0 and 1, and were then fed into the deep-learning models.

Results

Data Preprocessing

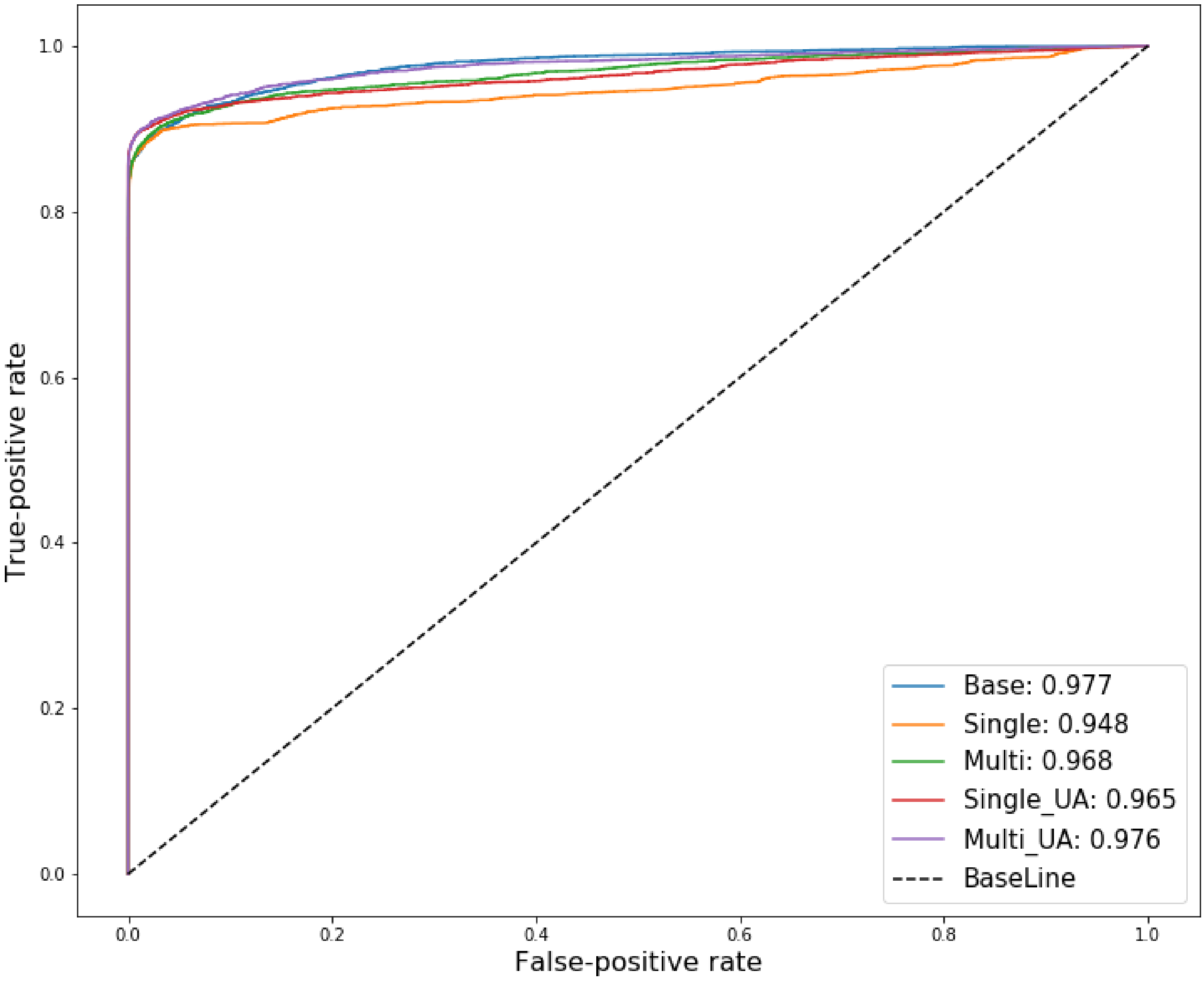
There were 238 variables with consistent labels in the 8-year cumulative dataset. Only 128 variables were selected by preprocessing. There were 22 variables with no change in value, 84 variables with more than 50% missing values, and 4 variables containing identification information that were excluded from the analysis. The total number of observations (ie, the number of subjects) was 33,065, with an average age of 48.89 years and with men accounting for 40.41% (n=13,361) of the sample. Only 6 variables had no omissions, and the average missing rate of variables with omissions was 10.38%

Outcome

Prediction Performance

Figure 5 represents the results for the ROC test and area under the curve (AUC) values of the five models. The results are based on the combined sets of predicted probabilities of 10 test sets of each model. According to the AUC results, the accuracy of the five models in terms of sensitivity and specificity was excellent overall. The AUCs of the base model without the attention mechanism, the single attention model, multi-attention model, single attention model with UA, and multi-attention model with UA were 0.977, 0.948, 0.968, 0.965, and 0.976, respectively.

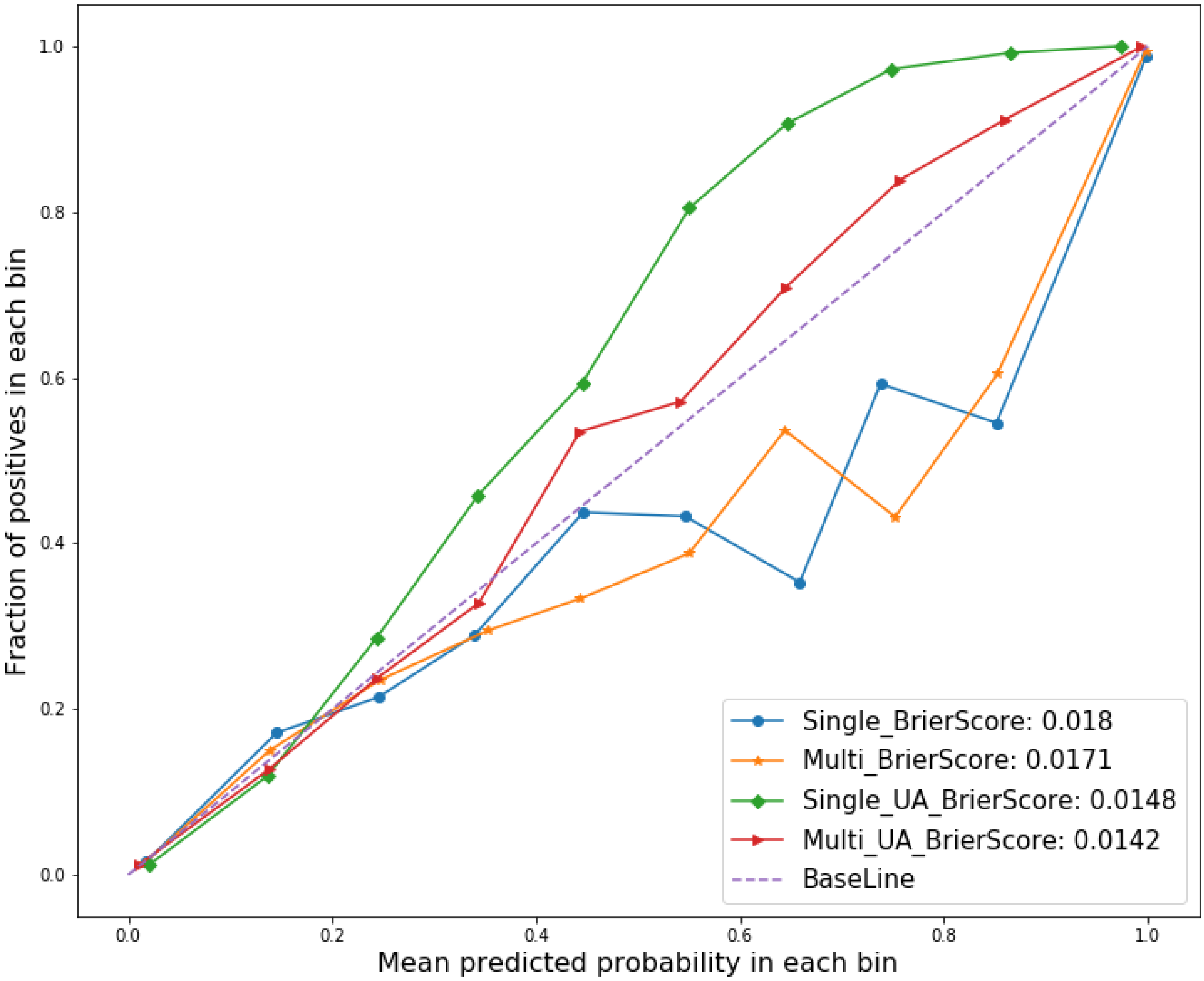
Figure 5. Receiver operating characteristic (ROC) test for the four models. All values estimated from 10 test sets are combined into one set for each model. The base indicates the deep-learning model without attention mechanisms. UA: uncertainty awareness.



Probability Reliability

Figure 6 shows the performance of probability reliability for the four models in the form of a reliability diagram. A characteristic of the UA models is that most fractions of positives were plotted above the diagonal. By contrast, models without UA showed more fractions of positives below the diagonal than the other models. The fraction of positives of the multi-attention with UA model was the closest to the diagonal. The Brier scores of the single attention, multi-attention, single attention with UA, and multi-attention with UA models were 0.018, 0.0171, 0.0148, and 0.0142, respectively.

Figure 6. Reliability diagrams for four models. The Brier score measures the overall reliability of probabilistic predictions. UA: uncertainty awareness.

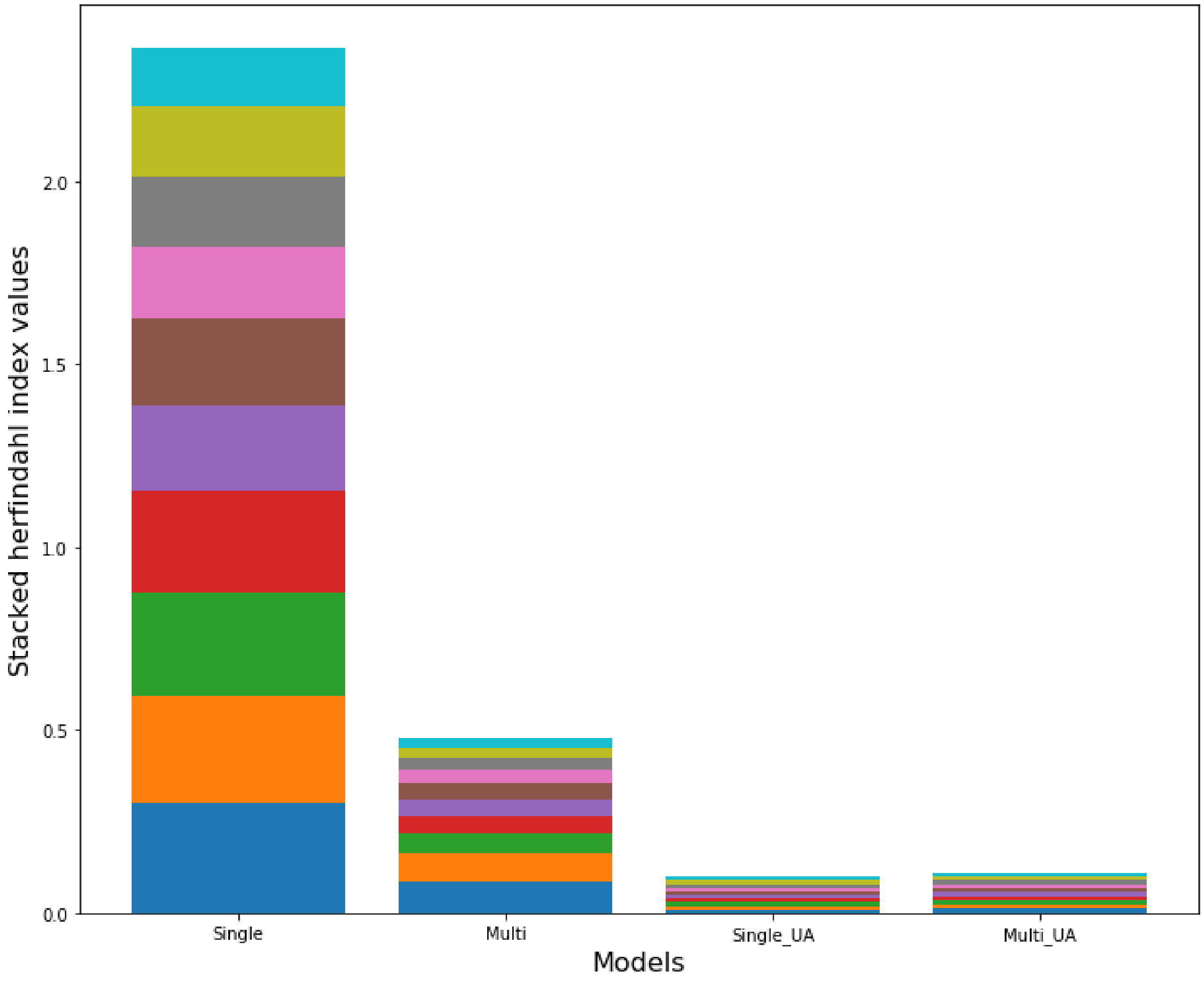


Attention

Concentration

Figure 7 shows stacked Herfindahl indices sorted by value size for the 10-fold sets in each model. The values for each fold are presented in Table A1 of Multimedia Appendix 1. In general, models without UA showed relatively large Herfindahl indices. The average Herfindahl index values for the single attention and multi-attention models were 0.236 and 0.048, respectively. However, models with UA had very small values, regardless of the degree of freedom. The average Herfindahl index values for the single attention with UA and multi-attention with UA models were 0.01 and 0.01, respectively. These results indicate that influence is more concentrated on several variables in models where uncertainty is not considered than in those where uncertainty is considered.

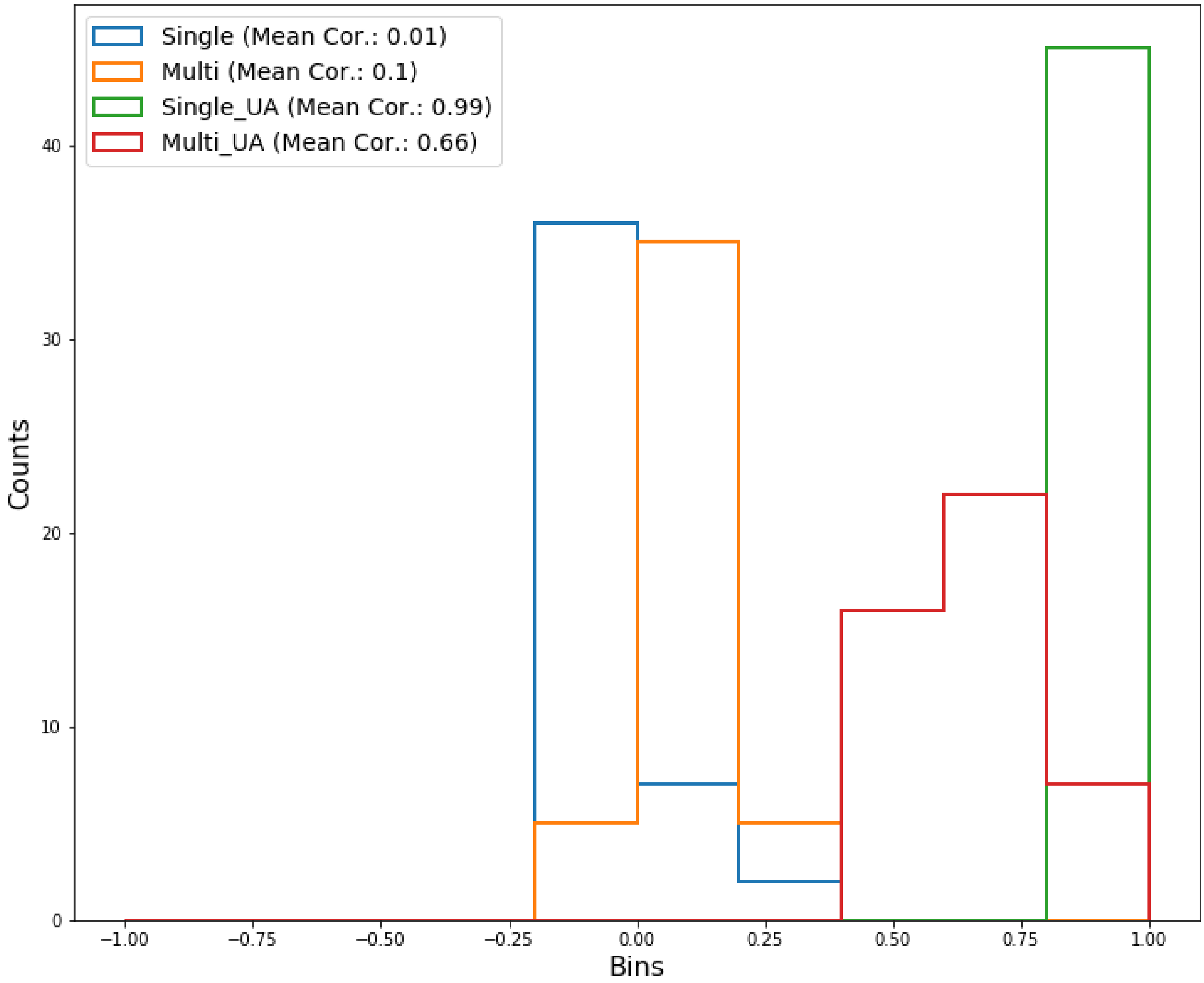
Figure 7. Herfindahl index values of 10-fold sets for each model. UA: uncertainty awareness.



Consistency

Figure 8 shows histograms for the 45 correlations ({[10×10]–10}/2) among the 10-fold sets for each model. In general, the correlations of the fold sets from the models with UA were higher than those of the models without UA. The average correlations of the fold sets from both models without UA were calculated to be close to zero (ie, 0.01 and 0.1). Moreover, the average correlations from the two models with UA were calculated as 0.99 and 0.66, respectively.

Figure 8. Histograms for correlations among 10-fold sets for each model. UA: uncertainty awareness.



Generalizability

Table 2 shows the results of the variable importance learned by the attention models and the effect sizes measured by conventional statistical methods. Definitions of each variable are provided in Table A2 of Multimedia Appendix 1. The top 5% (127×0.05=6) of variables, sorted by the magnitude of values obtained from each method, are reported. Since Cohen *d* from conventional methods may take on negative values, the absolute value was applied when sorting. Overall, models in which uncertainty was not considered were trained to have high attention values. Furthermore, variables such as “allownc” (whether to receive a basic living allowance) and “house” (whether to have own house) that bear little relation to health status were included in the results.

Table 2. Top 5% variable importance results estimated by different methods.

Variable[a]	Attention value[b]/Effect size[c]
Single attention model	
sm_presnt	0.0615
BD1	0.0519
pa_walk	0.0479
HE_Upro	0.0468
HE_alt	0.0430
Npins	0.0413
Multi-attention model	
HE_HB	0.041
Sex	0.033
Pa_walk	0.032
HE_HBsAg	0.027
Allownc	0.026
HE_sbp	0.026
Single attention model with UA[d]	
pa_walk	0.050
BH9_11	0.017
HE_THfh2	0.011
HE_THfh1	0.010
HE_THfh3	0.010
DI5_dg	0.010
Multi-attention model with UA	
pa_walk	0.050
HE_THfh2	0.019
BH9_11	0.015
HE_THfh1	0.013
HE_ast	0.010
house	0.010
Conventional statistics[e]	
age	1.536
HE_glu	1.214
HE_HbA1c	–1.014
Wt_pool_1	–0.516
Wt_itvex	–0.516
HE_Uglu	0.319

[a] See Table A2 in Multimedia Appendix 1 for variable label descriptions.

[b] Average from the 10 fold sets.

[c] Effect size is presented only for the conventional statistics.

[d] UA: uncertainty awareness.

[e] Null hypothesis of categorical variables=no relationships between diabetes and a categorical variable; null hypothesis of continuous variables=no difference in variables between the diabetes and no diabetes groups.

Table 3 shows the overall relationship between the effect sizes of variables and variable importance. Since attention values from both single attention and multi-attention models with UA had a high correlation (0.943), two regression models in which the two variables did not overlap were specified. The regression results showed no association between the variable importance from attention models and the effect size of variables from conventional methods.

Table 3. Regression analysis results for assessing an association between attention values and effect sizes.

Regression model variables[a]	Regression 1		Regression 2	
	Coefficient	$P > t$	Coefficient	$P > t$
Single attention	−0.696	0.635	−0.668	0.652
Multi-attention	1.897	0.487	2.019	0.476
Single attention with UA[b]	−1.351	0.797	—[c]	—
Multi-attention with UA	—	—	−1.541	0.772
Intercept	0.075	0.09	0.075	0.075

[a] The dependent variable is the absolute value of effect size, calculated by Cohen *d* for continuous variables and Cramer *V* for categorical variables. The total number of observations is equal to the number of variables.

[b] UA: uncertainty awareness.

[c] —: variable not included in the regression model.

Discussion

Principal Findings

Reliability

A difference in performance according to the degree of freedom was prominent in the probability reliability diagram (Figure 6). The fraction of positives located above the diagonal indicates that probabilities are predicted to be larger than expected, while the fraction of positives located below the diagonal means that probabilities are estimated to be smaller than expected [35,36]. In this regard, overall probabilities from the two attention models without UA tended to be underestimated, whereas the attention models with UA tended to overestimate probabilities. Although no clear causal relationship has been identified, several lines of empirical evidence suggest that the over- and underestimation is associated with data noise, estimation methods, and parameter settings [30,52-54].

Since the difference appeared to be based on the UA axis, the over- or underestimation tendency of the models may be related to UA. Furthermore, the Brier scores of the two models with UA were smaller than those of the two models without UA, indicating that models with UA tend to estimate more reliable probabilities than models without UA. These findings are consistent with the results of recent research that estimated reliable outcomes with an emphasis on UA [9,55-57]. Theoretically, the most probabilistic values are inferred from a distribution that takes means and variances into account under UA [9,55-57]. Thus, the awareness of uncertainty may bring reliability to the prediction results of deep-learning models, which are vulnerable to randomness during the learning process.

Consistency

UA produced noticeable differences in results consistency and the concentration of variable importance. Specifically, in UA models with low Herfindahl indices, variable importance appeared to be distributed over many variables in contrast to models that did not consider uncertainty (Figure 7). In addition, high correlations between 10-fold sets were found in the attention results from the UA models, whereas no correlations were found in the results from UA models (Figure 8). Furthermore, the attention values in the models with UA were generally smaller than those of models without UA.

These results suggest that the consistency of results from the UA models is high because the variable importance with overall low values is distributed evenly over most variables. This result is closely associated with the assumption that attention values were estimated based on a normal distribution within the cost function (see equation for the Kullback–Leibler divergence D_{KL} in Notation A2 of Multimedia Appendix 1). According to this equation, as both μ and σ^2 approach zero, model parameters for forming the normal distribution approximate the true theoretical distribution, indicating that the models are well learned [32,33]. Consequently, the overall attention values were small since the overall values of μ were small.

Spurious Correlations

As with conventional statistical methods, the attention models were unable to control spurious correlations during attention learning. Specifically, of the top 5% of variables obtained from conventional statistics, wt_pool_1 (interview weight combined years) and wt_itvex (interview weight for a single year) have little to do with health status (Table 2). These variables are weights for compensating errors due to differences in the number of households and populations between the sample design time and the survey time. In addition, the variables “allownc” (whether to receive basic living allowance) and “house” (whether to have own house) were obtained from the attention models (Table 2). These results may suggest spurious correlations in the dataset itself [58]. In other words, these variables, with little relation to diabetes, have a relatively close relationship with diabetes only by chance.

Generalizability to Conventional Statistics

In terms of clinically relevant variables, no significant association between the results of conventional statistics and attention models was found. Specifically, the variables age, HE_glu (fasting blood sugar), HE_HbA1c (hemoglobin AIC), and HE_Uglu (urine glucose) selected by conventional methods are well known to have a direct association with diabetes [59-61]. In addition, several variables, including pa_walk (amount of walking) and BH9_11 (vaccination status against influenza virus), obtained by the attention models are less directly related to diabetes. These variables may represent behavioral characteristics of patients with diabetes who are trying to manage their health.

Furthermore, there was no intersection of variables selected by both attention models and conventional statistical methods (Table 2). In particular, HE_glu, which was intentionally used as an input variable for testing purposes, was not determined as a major variable in the attention mechanism models in contrast to the conventional statistical methods. Additionally, no variable was statistically significant in the regression analysis that evaluated the positive association between attention values and effect sizes (Table 3). Comprehensively, these results suggest that the variable importance obtained from attention mechanisms may not be generalized to the effect size of conventional statistics.

Lessons from the Findings

Hyperparameters

The model structure and weight of terms in the cost function are hyperparameters to be adjusted. In terms of the model structure, the degree of freedom of attention layers was evaluated by comparing two extreme cases of 1 attention layer and 20 attention layers. Although the size of attention layers does not make a significant difference, the results can be significantly different if the number of attention layers is different under other conditions.

Furthermore, by taking uncertainty into account in the models, a term (ie, the degree of similarity associated with the normal distribution) was added to the cost function. However, as discussed previously, this term may interfere with the assignment of great importance values to variables by making all μ values small. To alleviate this issue, the weight of the term may be lowered, so that the term is less reflected during model training [9,29,33].

However, hyperparameter tuning is not conducted based on a theoretical basis but rather on a heuristic basis. In other words, there is no standard of how many attention layers should be specified and how much the weight should be adjusted for better results. If the goal of building models aims to maximize accuracy, various hyperparameter settings can be tested in the direction of increasing model accuracy. However, there is no clear criterion to maximize the performance of interpretability. In other words, although various hyperparameter settings are tested, finding the best-optimized hyperparameter setting based on the statistical point of view is challenging. Therefore, the variable importance should be understood in a limited way only within the framework of this experiment.

Potential Limitations of Interpretability

There was no significant association between the variable importance obtained from the attention mechanism and the effect size obtained from conventional statistics. One of the most probable reasons for this result is that the assumption of the association among input variables is different between conventional methods and deep-learning algorithms. Specifically, conventional statistical methodologies such as linear regression analysis and ANOVA basically estimate effect sizes based on the assumption of independence between input variables [62,63]. Thus, if a particular input variable has nothing to do with an outcome variable, the variable has little effect on the outcome. In contrast, neural network–based algorithms, including deep learning, infer outcome variables by taking into account the dependencies between the input variables [41,42]. Therefore, a variable that is not directly related to an outcome variable but is associated with others that are related to the outcome variable may have a somewhat greater effect on the outcome variable. Owing to these differences, attention results must not be considered to have similar meanings and tendency to the variable effect size from conventional methodologies.

Recent new technologies such as sensors (ie, wearable devices or facilities in operating rooms) have produced new types of data. Since the associations between variables have not yet been fully explored, relying solely on attention mechanisms may lead to a false judgment that variables that have minimal association with the outcome variable are important. Hence, it is advisable to consider the results of attention and conventional statistics together.

Furthermore, in situations where there is a spurious correlation, neither method provides good explanatory power. Spurious correlations can only be eliminated through data preprocessing based on domain knowledge. Hence, care must be taken when implementing both attention models and conventional statistical methods in environments with manifold variables that cannot be preprocessed (ie, included or excluded) using definite knowledge. In particular, finding new features using attention mechanisms may not be adequate in environments where the data are susceptible to spurious correlations owing to a large number of variables but few observations such as in the field of genetic engineering [64,65]. In this case, it may be appropriate to employ results of attention mechanisms for reaffirming existing findings in previous research or supporting informed knowledge.

Future Direction for Medical Informatics

The results of this study provide several points of guidance for future research in the medical field. First, more empirical evidence should be secured based on various structures in terms of the degree of freedom. It may be desirable to test what attention results are produced when different values of degree of freedom are employed. Particularly, given that the medical field has various data types such as images, natural languages, and numerical values, attention results should be assessed according to the degree of freedom with consideration of the data characteristics [66-69].

Second, attention models with more sophisticated UA should be tested. In this study, the model outcome variable was assumed to depend on the distribution of the attention layer; that is, $P(diabetes|z)$. However, current state-of-the-art Bayesian estimation assumes that the model outcomes depend on all network weights and data; that is, $P(outcome|z, weight, data)$ [9,29]. Thus, it is necessary to evaluate how the variable importance is formed when more up-to-date estimation methods are applied.

Third, more research that strictly evaluates variable importance based on attention mechanisms over diverse disease domains is needed. As found in this study, attention has its limitations in terms of generalizability to conventional statistics and control of spurious correlations. However, since this case study was conducted with a single cohort of Korean patients with diabetes, more empirical evidence from various cohorts or diseases should be tested to confirm that attention mechanisms may not provide any significant meaning. Importantly, for elaborate empirical research, a greater in-depth understanding of the association between covariates and health outcomes is needed. Hence, more domain experts on a specific disease along with data scientists should be actively involved in these studies.

Fourth, methods for controlling the distribution of variable importance should be studied. As revealed in this analysis, the variable importance can be distributed over many variables or concentrated on a few variables depending on the model structure (Figure 7). When examining the overall relationships between covariates and health outcomes such as a comprehensive review of national health status [45-49], it may be desirable to detect many potentially important variables. By contrast, when the relationship between a small number of key variables and outcome is important, such as in the generation of targeted therapy [70,71], the importance should be focused on a few variables. However, to the best of our knowledge, most existing attention studies have not considered the control of the variable importance distribution [8,10-12,24,25,32,33,66,68]. Therefore, more studies on this subject are needed.

Limitations

There are several limitations to be aware of when assessing the academic value of this study. First, well-behaved data with excellent predictive performance owing to the data characteristics were employed for the analysis. For this reason, the overall AUC performance (see ROC test in Figure 5) might have been good for all approaches (ie, the degree of freedom and UA). When the attention mechanisms are applied to ill-behaved data without manipulation, such as the intentional use of a variable HE_glu as an input variable, the model accuracy may be reduced. If the accuracy of the model is moderate and domain expertise exists for the disease, it is still advisable to attempt a variable importance interpretation. However, if the model accuracy becomes too poor, it may not be worthwhile to interpret the variable importance. Furthermore, categorical variables of both nominal and ordinal types were integerized, and missing values were encoded as the extreme value in this study. Although this operationalization can be efficient in deep-learning algorithms that can learn nonlinear relationships, it is not a robust approach. Thus, it is necessary to identify problems with the approach and to discuss how to deal with them when ill-behaved data with robust operationalization are employed. Furthermore, since data from a single cohort were used, the results of this study, which point out the limitations of the interpretable power of attention mechanisms, should not be generalized. Rather, it should be recognized that accuracy performance and interpretable power may vary depending on the modeling approaches and data.

Second, this study does not guarantee that state-of-the-art estimation methods for UA were applied. Specifically, the models' outcomes do not depend on network weights. In addition, research on estimation methodologies in deep learning is in progress, and therefore new methodologies are still being developed. Accordingly, the value of this study lies in the framework proposals that suggest the research direction of attention modeling rather than in the details of attention estimation methods.

Third, the design of weights of each local attention layer is not as sophisticated as the design of local attention layers (Figure 4). Specifically, uncertainty considerations are not assumed in the weights layer. Moreover, this layer does not have to be dependent on the local attention layer. In other words, the weights layer may be designed as an independent layer that does not come from the local attention layer. We plan to perform various investigations in this area.

Conclusions

Attention mechanisms have the potential to make a significant contribution to the medical field, where explanatory power is important, by overcoming the limitations of the noninterpretability of deep-learning algorithms. However, potential problems that may arise when attention mechanisms are applied in practice have not been well studied. Thus, we hope that this study will serve as a cornerstone to raise potential issues, and that many similar studies will be conducted in the future. The cohesive awareness of potential problems arising from attention mechanisms in the field of application will provide theoretical researchers with new goals for problem-solving.

Acknowledgments

This study was supported by a grant (2010380-1) from the National Cancer Center of Korea.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Notations, global attention inference procedure, Herfindahl index values, variable label descriptions, and reparameterization trick concept description.
[PDF File (Adobe PDF File), 525 KB - jmir_v22i12e18418_app1.pdf]

Multimedia Appendix 2
Codes and algorithms for inferring outcomes.
[PDF File (Adobe PDF File), 381 KB - jmir_v22i12e18418_app2.pdf]

References

1. Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, Oskotsky B, et al. Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. JAMA Netw Open 2019 Mar 01;2(3):e190606 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0606] [Medline: 30874779]
2. Shamai G, Binenbaum Y, Slossberg R, Duek I, Gil Z, Kimmel R. Artificial Intelligence Algorithms to Assess Hormonal Status From Tissue Microarrays in Patients With Breast Cancer. JAMA Netw Open 2019 Jul 03;2(7):e197700 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.7700] [Medline: 31348505]
3. Montavon G, Samek W, Müller K. Methods for interpreting and understanding deep neural networks. Dig Sign Process 2018 Feb;73:1-15. [doi: 10.1016/j.dsp.2017.10.011]
4. Zhang Z, Beck MW, Winkler DA, Huang B, Sibanda W, Goyal H, written on behalf of AME Big-Data Clinical Trial Collaborative Group. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. Ann Transl Med 2018 Jun;6(11):216. [doi: 10.21037/atm.2018.05.32] [Medline: 30023379]
5. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. 2015 Presented at: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining; August 2015; Sydney, Australia p. 1721-1730. [doi: 10.1145/2783258.2788613]
6. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. 2017 Presented at: Advances in neural information processing systems; December 2017; Long Beach, CA p. 4765-4774.
7. Ribeiro M, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; August 2016; New York, NY p. 1135-1144.
8. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. 2017 Presented at: Advances in Neural Information Processing Systems; December 2017; Long Beach, CA p. 5998-6008.
9. Heo J, Lee H, Kim S, Lee J, Kim K, Yang E, et al. Uncertainty-aware attention for reliable interpretation and prediction. 2018 Presented at: Advances in Neural Information Processing Systems; December 2018; Montréal, Canada p. 909-918.
10. Tomita N, Abdollahi B, Wei J, Ren B, Suriawinata A, Hassanpour S. Attention-Based Deep Neural Networks for Detection of Cancerous and Precancerous Esophagus Tissue on Histopathological Slides. JAMA Netw Open 2019 Nov 01;2(11):e1914645 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.14645] [Medline: 31693124]
11. Hwang EJ, Park S, Jin K, Kim JI, Choi SY, Lee JH, DLAD Development Evaluation Group. Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. JAMA Netw Open 2019 Mar 01;2(3):e191095 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.1095] [Medline: 30901052]
12. Kaji DA, Zech JR, Kim JS, Cho SK, Dangayach NS, Costa AB, et al. An attention based deep learning model of clinical events in the intensive care unit. PLoS One 2019;14(2):e0211057 [FREE Full text] [doi: 10.1371/journal.pone.0211057] [Medline: 30759094]
13. Li L, Zhao J, Hou L, Zhai Y, Shi J, Cui F. An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records. BMC Med Inform Decis Mak 2019 Dec 05;19(Suppl 5):235 [FREE Full text] [doi: 10.1186/s12911-019-0933-6] [Medline: 31801540]
14. You R, Liu Y, Mamitsuka H, Zhu S. BERTMeSH: Deep Contextual Representation Learning for Large-scale High-performance MeSH Indexing with Full Text. Bioinformatics 2020 Sep 25:btaa837. [doi: 10.1093/bioinformatics/btaa837] [Medline: 32976559]
15. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: Learning to leverage salient regions in medical images. Med Image Anal 2019 Apr;53:197-207 [FREE Full text] [doi: 10.1016/j.media.2019.01.012] [Medline: 30802813]
16. Oktay O, Schlemper J, Folgoc L, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. ArXiv 2018:1804.03999 [FREE Full text]
17. Hasan S, Ling Y, Liu J, Sreenivasan R, Anand S, Arora T, et al. Attention-based medical caption generation with image modality classification and clinical concept mapping. 2018 Presented at: International Conference of the Cross-Language Evaluation Forum for European Languages; September 2018; Avignon, France p. 224-230.
18. Hahn M. Theoretical Limitations of Self-Attention in Neural Sequence Models. Trans Assoc Comput Ling 2020 Jul;8:156-171. [doi: 10.1162/tacl_a_00306]

19. Ghorbani A, Abid A, Zou J. Interpretation of Neural Networks Is Fragile. 2019 Jul 17 Presented at: Proceedings of the AAAI Conference on Artificial Intelligence; February 7-12, 2020; New York, NY p. 3681-3688. [doi: 10.1609/aaai.v33i01.33013681]
20. Feldman R, Sanger J. The text mining handbook: Advanced approaches in analyzing unstructured data. Cambridge, UK: Cambridge University Press; 2007.
21. Ryan M. Deep learning with structured data. Shelter Island, NY: Manning Publications; 2020.
22. Li Y, Rao S, Solares J, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: Transformer for Electronic Health Records. Sci Rep 2020 Apr 28;10(1):7155. [doi: 10.1038/s41598-020-62922-y] [Medline: 32346050]
23. Choi E, Bahadori M, Sun J, Kulas J, Schuetz A, Stewart W. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. 2016 Presented at: Advances in Neural Information Processing Systems; December 2016; Barcelona, Spain p. 3504-3512.
24. Chorowski J, Bahdanau D, Serdyuk D, Cho K, Bengio Y. Attention-based models for speech recognition. 2015 Presented at: Advances in Neural Information Processing Systems; December 2015; Quebec, Canada p. 577-585.
25. Stollenga M, Masci J, Gomez F, Schmidhuber J. Deep networks with internal selective attention through feedback connections. 2014 Presented at: Advances in Neural Information Processing Systems; December 2014; Quebec, Canada p. 3545-3553.
26. Devlin J, Chang M, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv 2018:1810.04805 [FREE Full text]
27. Song H, Rajan D, Thiagarajan J, Spanias A. Attend and diagnose: Clinical time series analysis using attention models. arXiv 2018:preprint 1711.03905 [FREE Full text]
28. Xiong Y, Du B, Yan P. Reinforced Transformer for Medical Image Captioning. 2019 Presented at: International Workshop on Machine Learning in Medical Imaging; October 2019; Shenzhen, China p. 680. [doi: 10.1007/978-3-030-32692-0_77]
29. Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. 2016 Presented at: International conference on machine learning; June 2016; New York, NY p. 1050-1059.
30. Gal Y. Uncertainty in deep learning. University of Cambridge, PhD Thesis. 2016. URL: http://mlg.eng.cam.ac.uk/yarin/thesis/thesis.pdf [accessed 2020-12-07]
31. Kläs M, Vollmer A. Uncertainty in machine learning applications: A practice-driven classification of uncertainty. 2018 Presented at: International Conference on Computer Safety, Reliability, and Security; September 2018; Västerås, Sweden p. 431-438.
32. Ehsan AM, Dick A, van DHA. Infinite variational autoencoder for semi-supervised learning. 2017 Presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; July 2017; Honolulu, USA p. 5888-5897. [doi: 10.1109/cvpr.2017.90]
33. Kingma D, Welling M. Auto-encoding variational Bayes. arXiv 2013:1312.6114 [FREE Full text]
34. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982 Apr;143(1):29-36. [doi: 10.1148/radiology.143.1.7063747] [Medline: 7063747]
35. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. 2005 Presented at: Proceedings of the 22nd international conference on machine learning; August 2005; New York, NY p. 625-632. [doi: 10.1145/1102351.1102430]
36. Bröcker J, Smith L. Increasing the reliability of reliability diagrams. Weather Forecast 2007;22(3):651-661 [FREE Full text]
37. Benedetti R. Scoring rules for forecast verification. Month Weather Rev 2010;138(1):203-211 [FREE Full text]
38. Kwoka JJ. The Herfindahl index in theory and practice. Antitrust Bull 1985;30:915.
39. Cramér H. Mathematical Methods of Statistics (PMS-9). Princeton, NJ: Princeton University Press; 2016.
40. Rosenthal R, Cooper H, Hedges L. Parametric measures of effect size. In: Cooper H, Hedges LV, editors. The handbook of research synthesis. New York, NY: Russel Sage Foundation; 1994:231-244.
41. Brownlee J. Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras. Vermont, Victoria, Australia: Machine Learning Mastery; 2016.
42. Chollet F. Deep learning with Python. Shelter Island, NY: Manning Publications Co; 2018.
43. Kingma D, Salimans T, Welling M. Variational dropout and the local reparameterization trick. 2015 Presented at: Advances in neural information processing systems; December 2015; Quebec, Canada p. 2575-2583.
44. Kweon S, Kim Y, Jang M, Kim Y, Kim K, Choi S, et al. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). Int J Epidemiol 2014 Feb;43(1):69-77 [FREE Full text] [doi: 10.1093/ije/dyt228] [Medline: 24585853]
45. Moon S. Low skeletal muscle mass is associated with insulin resistance, diabetes, and metabolic syndrome in the Korean population: the Korea National Health and Nutrition Examination Survey (KNHANES) 2009-2010. Endocr J 2014;61(1):61-70 [FREE Full text] [doi: 10.1507/endocrj.ej13-0244] [Medline: 24088600]
46. Jee D, Lee WK, Kang S. Prevalence and risk factors for diabetic retinopathy: the Korea National Health and Nutrition Examination Survey 2008-2011. Invest Ophthalmol Vis Sci 2013 Oct 17;54(10):6827-6833. [doi: 10.1167/iovs.13-12654] [Medline: 24065813]

47. Choi YJ, Lee MS, An SY, Kim TH, Han SJ, Kim HJ, et al. The Relationship between Diabetes Mellitus and Health-Related Quality of Life in Korean Adults: The Fourth Korea National Health and Nutrition Examination Survey (2007-2009). Diabetes Metab J 2011 Dec;35(6):587-594 [FREE Full text] [doi: 10.4093/dmj.2011.35.6.587] [Medline: 22247901]
48. Hwang J, Shon C. Relationship between socioeconomic status and type 2 diabetes: results from Korea National Health and Nutrition Examination Survey (KNHANES) 2010-2012. BMJ Open 2014 Aug 19;4(8):e005710 [FREE Full text] [doi: 10.1136/bmjopen-2014-005710] [Medline: 25138810]
49. Ahn JH, Yu JH, Ko S, Kwon H, Kim DJ, Kim JH, Taskforce Team of Diabetes Fact Sheet of the Korean Diabetes Association. Prevalence and determinants of diabetic nephropathy in Korea: Korea national health and nutrition examination survey. Diabetes Metab J 2014 Apr;38(2):109-119 [FREE Full text] [doi: 10.4093/dmj.2014.38.2.109] [Medline: 24851205]
50. Au TC. Random forests, decision trees, and categorical predictors: the "Absent Levels" Problem. J Machine Learn Res 2018;19(1):1737-1766 [FREE Full text]
51. Maass P. Deep learning for trivial inverse problems. In: Compressed Sensing and its Applications. Cham: Birkhäuser; 2019:195-209.
52. Thrun S, Schwartz A. Issues in using function approximation for reinforcement learning. 1993 Presented at: Proceedings of the 1993 Connectionist Models Summer School; December 1993; Hillsdale, NJ.
53. Hayward RA, Heisler M, Adams J, Dudley RA, Hofer TP. Overestimating outcome rates: statistical estimation when reliability is suboptimal. Health Serv Res 2007 Aug;42(4):1718-1738 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00661.x] [Medline: 17610445]
54. Dehmer M, Basak S. Statistical and Machine Learning Approaches for Network Analysis. Hoboken, NJ: Wiley Online Library; 2012.
55. Sander J, de VB, Wolterink J, Išgum I. Towards increased trustworthiness of deep learning segmentation methods on cardiac MRI. 2019 Presented at: Medical Imaging 2019: Image Processing; March 2019; San Diego, CA p. 1094919.
56. Kuleshov V, Liang P. Calibrated structured prediction. 2015 Presented at: Advances in Neural Information Processing Systems; December 2015; Quebec, Canada p. 3474-3482.
57. Seo S, Seo P, Han B. Learning for single-shot confidence calibration in deep neural networks through stochastic inferences. 2019 Presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; June 2019; Long Beach, CA p. 9030-9038.
58. Simon HA. Spurious Correlation: A Causal Interpretation. J Am Stat Assoc 1954 Sep;49(267):467. [doi: 10.2307/2281124]
59. Kirkman MS, Briscoe VJ, Clark N, Florez H, Haas LB, Halter JB, et al. Diabetes in older adults. Diabetes Care 2012 Dec;35(12):2650-2664 [FREE Full text] [doi: 10.2337/dc12-1801] [Medline: 23100048]
60. Koenig RJ, Cerami A. Hemoglobin A Ic and diabetes mellitus. Annu Rev Med 1980;31(1):29-34. [doi: 10.1146/annurev.me.31.020180.000333] [Medline: 6994614]
61. Lev-Ran A. Glycohemoglobin. Arch Intern Med 1981 May 01;141(6):747. [doi: 10.1001/archinte.1981.00340060055013]
62. Heiss F. Using R for introductory econometrics. South Carolina: Createspace Independent Publishing Platform; 2016.
63. Cannon A, Cobb G, Hartlaub B, Legler J, Lock R, Moore T, et al. Stat2: modeling with regression ANOVA. Second edition. New York, NY: WH Freeman; 2018.
64. Lee KE, Sha N, Dougherty ER, Vannucci M, Mallick BK. Gene selection: a Bayesian variable selection approach. Bioinformatics 2003 Jan;19(1):90-97. [doi: 10.1093/bioinformatics/19.1.90] [Medline: 12499298]
65. Wimmer V, Lehermeier C, Albrecht T, Auinger H, Wang Y, Schön CC. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. Genetics 2013 Oct;195(2):573-587 [FREE Full text] [doi: 10.1534/genetics.113.150078] [Medline: 23934883]
66. Wang H, Wang Y, Liang C, Li Y. Assessment of Deep Learning Using Nonimaging Information and Sequential Medical Records to Develop a Prediction Model for Nonmelanoma Skin Cancer. JAMA Dermatol 2019 Sep 04;155(11):1277-1283 [FREE Full text] [doi: 10.1001/jamadermatol.2019.2335] [Medline: 31483437]
67. Beaulieu-Jones B, Finlayson SG, Chivers C, Chen I, McDermott M, Kandola J, et al. Trends and Focus of Machine Learning Applications for Health Research. JAMA Netw Open 2019 Oct 02;2(10):e1914051 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.14051] [Medline: 31651969]
68. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, the CAMELYON16 Consortium, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA 2017 Dec 12;318(22):2199-2210 [FREE Full text] [doi: 10.1001/jama.2017.14585] [Medline: 29234806]
69. Taggart M, Chapman WW, Steinberg BA, Ruckel S, Pregenzer-Wenzler A, Du Y, et al. Comparison of 2 Natural Language Processing Methods for Identification of Bleeding Among Critically Ill Patients. JAMA Netw Open 2018 Oct 05;1(6):e183451 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.3451] [Medline: 30646240]
70. Jerby-Arnon L, Pfetzer N, Waldman Y, McGarry L, James D, Shanks E, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell 2014 Aug 28;158(5):1199-1209 [FREE Full text] [doi: 10.1016/j.cell.2014.07.027] [Medline: 25171417]
71. Hyman DM, Taylor BS, Baselga J. Implementing Genome-Driven Oncology. Cell 2017 Feb 09;168(4):584-599 [FREE Full text] [doi: 10.1016/j.cell.2016.12.015] [Medline: 28187282]

Abbreviations

ANOVA: analysis of variance
AUC: area under the curve
LIME: Local Interpretable Model-agnostic Explanations
ROC: receiver operating characteristic
SHAP: Shapley Additive Explanations
UA: uncertainty awareness

Edited by G Eysenbach, R Kukafka; submitted 25.02.20; peer-reviewed by SY Shin, M Feng, S Azimi; comments to author 07.08.20; revised version received 13.09.20; accepted 11.11.20; published 16.12.20.

Please cite as:
Kim J, Lee S, Hwang E, Ryu KS, Jeong H, Lee JW, Hwangbo Y, Choi KS, Cha HS
Limitations of Deep Learning Attention Mechanisms in Clinical Research: Empirical Case Study Based on the Korean Diabetic Disease Setting
J Med Internet Res 2020;22(12):e18418
URL: http://www.jmir.org/2020/12/e18418/
doi:10.2196/18418
PMID:33325832

Original Paper

An Artificial Intelligence Model to Predict the Mortality of COVID-19 Patients at Hospital Admission Time Using Routine Blood Samples: Development and Validation of an Ensemble Model

Hoon Ko[1*], MSc; Heewon Chung[1*], MSc; Wu Seong Kang[2*], MD, PhD; Chul Park[3*], MD; Do Wan Kim[4], MD; Seong Eun Kim[5], MD; Chi Ryang Chung[6], MD, PhD; Ryoung Eun Ko[6], MD, PhD; Hooseok Lee[1], MSc; Jae Ho Seo[7], PhD; Tae-Young Choi[8], PhD; Rafael Jaimes[9], PhD; Kyung Won Kim[10], MD; Jinseok Lee[1], PhD

[1]Biomedical Engineering, Wonkwang University, Iksan, Republic of Korea
[2]Department of Trauma Surgery, Wonkwang University Hospital, Iksan, Republic of Korea
[3]Department of Internal Medicine, Wonkwang University Hospital, Iksan, Republic of Korea
[4]Department of Thoracic and Cardiovascular Surgery, Chonnam National University Medical School, Gwangju, Republic of Korea
[5]Department of Internal Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea
[6]Department of Critical Care Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[7]Department of Biochemistry, Wonkwang University School of Medicine, Iksan, Republic of Korea
[8]Department of Pathology, Wonkwang University School of Medicine, Iksan, Republic of Korea
[9]Biotechnology and Human Systems, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, United States
[10]Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Jinseok Lee, PhD
Biomedical Engineering
Wonkwang University
Iksan Daero
Iksan, 54538
Republic of Korea
Phone: 82 1638506970
Email: gonasago@gmail.com

Abstract

Background: COVID-19, which is accompanied by acute respiratory distress, multiple organ failure, and death, has spread worldwide much faster than previously thought. However, at present, it has limited treatments.

Objective: To overcome this issue, we developed an artificial intelligence (AI) model of COVID-19, named EDRnet (ensemble learning model based on deep neural network and random forest models), to predict in-hospital mortality using a routine blood sample at the time of hospital admission.

Methods: We selected 28 blood biomarkers and used the age and gender information of patients as model inputs. To improve the mortality prediction, we adopted an ensemble approach combining deep neural network and random forest models. We trained our model with a database of blood samples from 361 COVID-19 patients in Wuhan, China, and applied it to 106 COVID-19 patients in three Korean medical institutions.

Results: In the testing data sets, EDRnet provided high sensitivity (100%), specificity (91%), and accuracy (92%). To extend the number of patient data points, we developed a web application (BeatCOVID19) where anyone can access the model to predict mortality and can register his or her own blood laboratory results.

Conclusions: Our new AI model, EDRnet, accurately predicts the mortality rate for COVID-19. It is publicly available and aims to help health care providers fight COVID-19 and improve patients' outcomes.

(J Med Internet Res 2020;22(12):e25442) doi:10.2196/25442

KEYWORDS

COVID-19; artificial intelligence; blood samples; mortality prediction

Introduction

COVID-19 is a highly contagious infection caused by SARS-CoV2. In severe cases, COVID-19 causes acute respiratory distress, multiple organ failure, and, eventually, death [1]. As of November 2020, COVID-19 cases and deaths are approaching 60 million and 1.5 million, respectively, worldwide.

In a pandemic situation, the most important issue in the management of patients diagnosed with COVID-19 is to select patients at risk of high mortality in the early period of disease and to provide appropriate treatments [2]. Particularly, the condition of patients at high risk can rapidly deteriorate. Some papers reported that deceased COVID-19 patients initially had mild symptoms but suddenly transitioned to a critical stage, leading to death [3-5]. In Italy, 75% of deceased patients showed mild symptoms, such as fever, dyspnea, and cough, at admission to the hospital [1]. Thus, the development of a prognostic model to predict mortality as early as possible is very critical.

In this pandemic crisis, the shortage of resources and medical staff causes big problems in the health care system. Accordingly, artificial intelligence (AI) can aid in the management of COVID-19 patients. A recent research study has developed an AI prediction model of mortality based on blood test results [6]. In this study, Yan et al initially considered 73 blood-borne markers for the mortality prediction model; finally, three blood biomarkers were selected, including lactate dehydrogenase (LDH), lymphocyte, and high-sensitivity C-reactive protein (hs-CRP). This model predicted mortality with 90% accuracy based on a decision tree using an XGBoost classifier [7] to analyze feature importance.

However, Yan et al's study has drawbacks. First, the three biomarkers derived from the XGBoost-based feature selection may not be the best choices. Feature importance provides a score indicating how each feature contributes in the construction of decision trees within the model. However, due to the stochastic nature of machine learning algorithms, each feature's importance score may vary. Moreover, in decision tree algorithms, such as an XGBoost and a random forest (RF), when multiple features have the same gain during the split, a branch in a tree is made by randomly selecting features among them. Second, numerous studies have shown that the disease progression of COVID-19 is not only associated with LDH [2,8-11], lymphocyte [12,13], and hs-CRP [2,10,14-17] but also with other blood-based biomarkers, such as neutrophil counts [16,18,19], albumin [18,20,21], and prothrombin activity [18,22-24]. In our study, we developed an AI model using 28 biomarkers for predicting the mortality of COVID-19 patients. Third, the three biomarker-based AI models [6] predicted mortality 10 days before a patient's recovery or death. These limitations show that the model may not work for COVID-19 patients who have just been diagnosed and hospitalized.

Therefore, in this study, we aimed to develop an AI model based on a blood test for mortality prediction at the early stage of hospital admission. We deployed the developed AI model on a public website so that all patients and medical staff could predict mortality using individual patient blood test results.

Methods

Data Sets

This study was approved by Wonkwang University Hospital (WKUH), Chonnam National University Hospital (CNUH), and Samsung Medical Center (SMC) in Korea. Informed consent was waived. For training data, we used the blood test results obtained from 375 COVID-19 patients collected between January 10, 2020, and February 24, 2020, in Tongji Hospital, Wuhan, China [6]. Of these, 14 patients without a blood test within 1 day after the hospital admission were excluded, and 361 patients—212 males (58.7%) and 149 females (41.3%); mean age 58.9 years (SD 16.5)—were included. As presented in Multimedia Appendix 1, the training data set of 361 patients included the admission date and time, discharge date and time, age, gender, mortality outcome, and results of blood tests obtained within 24 hours after hospital admission. For testing data, we collected medical records on COVID-19 patients (N=106) from three medical institutions: CNUH (85/106, 80.2%), WKUH (11/106, 10.4%), and SMC (10/106, 9.4%). The blood laboratory results from these 106 COVID-19 patients were collected between February 2020 and July 2020. Similar to the training data, we used the blood test data obtained within 24 hours after hospital admission (see Multimedia Appendix 2). For summarizing the statistics of the training and testing data sets, the patients were classified into a survivor group and a deceased group in the training and testing data sets. The number of blood tests differed across patients and institutions. The mean numbers of blood tests per patient were 61.21 (range 24-73) in the training data set and 35.36 (range 30-55) in the testing data set. The mean numbers of hospitalization days were 13.82 (survivor group) and 8.16 (deceased group) in the training data set and 18.21 (survivor group) and 17.98 (deceased group) in the testing data set (see Table 1).

Table 1. Statistical summary of the training and testing data sets.

Patient data	Training data set (N=361)	Testing data set (N=106)			
	Tongji Hospital	CNUH[a] (n=85)	WKUH[b] (n=11)	SMC[c] (n=10)	Total (N=106)
Number of patients, n (%)					
Total	361 (100)	85 (100)	11 (100)	10 (100)	106 (100)
Survived	195 (54.0)	85 (100)	9 (82)	10 (100)	104 (98.1)
Deceased	212 (58.7)	0 (0)	2 (18)	0 (0)	2 (1.9)
Gender, n (%)					
Male	212 (58.7)	34 (40)	5 (45)	3 (30)	42 (39.6)
Female	149 (41.3)	51 (60)	6 (55)	7 (70)	64 (60.4)
Number of hospitalization days, mean (SD)					
Survived	13.82 (6.38)	15.06 (7.90)	28.16 (11.13)	30.95 (23.03)	18.21 (11.46)
Deceased	8.16 (7.38)	N/A[d]	17.98 (11.83)	N/A	17.98 (11.83)
Age in years, mean (SD)	58.91 (16.49)	44.14 (21.81)	56.27 (23.00)	58.20 (21.05)	46.73 (22.28)
Number of blood biomarkers collected					
Min-max	24-73	32-55	30-52	30-40	30-55
Mean (SD)	61.21 (6.92)	36.89 (4.11)	35.00 (5.31)	34.20 (3.16)	35.36 (4.19)

[a] CNUH: Chonnam National University Hospital.

[b] WKUH: Wonkwang University Hospital.

[c] SMC: Samsung Medical Center.

[d] N/A: not applicable; there were no deceased patients in the testing data set at this institution.

Feature Selection

Given the total 73 blood biomarkers from the training data, we performed an analysis of variance (ANOVA), which uses an F test to check for any significant difference between the two groups (ie, deceased vs survivor) according to each blood biomarker. For the feature selection, we also considered the available data rate (ADR), which refers to how much blood biomarker data were available for training the AI model. This is calculated as

where $N_{patients}$ is the total number of patients (N=361) and $N_{biomarker}$ is the number of patients having each of the specific biomarker data.

Based on the ANOVA, we first selected the top 32 biomarkers corresponding to P values less than 10^{-5}. Subsequently, we excluded four biomarkers with ADR values of less than 90%. Table 2 summarizes the final selection of 28 biomarkers with the corresponding ANOVA P values and ADR values. The ANOVA P values and ADR values for all 73 biomarkers in the training data set are summarized in Multimedia Appendix 3, Table S1. The sample distributions of the selected 28 biomarkers in the survivor and deceased groups are presented in Multimedia Appendix 3, Figure S1.

Table 2. List of 28 blood biomarkers selected for the artificial intelligence model training.

Biomarker index No.	Blood biomarker	ANOVA[a] *P* value	ADR[b], %
1	Lymphocytes	2.44×10^{-46}	96.95
2	Neutrophils	5.65×10^{-43}	96.68
3	Albumin	2.90×10^{-37}	96.12
4	Lactate dehydrogenase	4.18×10^{-36}	96.12
5	Neutrophil count	3.54×10^{-35}	96.68
6	Hypersensitive C-reactive protein	8.38×10^{-35}	94.74
7	Prothrombin activity	3.20×10^{-26}	94.18
8	Calcium	2.24×10^{-19}	95.29
9	Urea	3.29×10^{-17}	96.12
10	Estimated glomerular filtration rate	5.05×10^{-17}	96.12
11	Monocytes	1.09×10^{-14}	96.95
12	Globulin	6.06×10^{-13}	96.12
13	Eosinophils	2.07×10^{-12}	96.68
14	Glucose	2.39×10^{-11}	93.63
15	Red blood cell distribution width (RDW)	8.43×10^{-10}	92.24
16	HCO_3^- (bicarbonate)	2.68×10^{-9}	96.12
17	RDW standard deviation	3.06×10^{-9}	92.24
18	Platelet count	1.46×10^{-8}	96.68
19	Mean platelet volume	1.92×10^{-7}	92.24
20	Platelet large-cell ratio	2.02×10^{-7}	92.24
21	Prothrombin time	3.42×10^{-7}	94.18
22	Total protein	5.29×10^{-7}	96.12
23	Platelet distribution width	6.98×10^{-7}	92.24
24	Aspartate aminotransferase	1.01×10^{-6}	96.12
25	Thrombocytocrit	1.49×10^{-6}	92.24
26	Eosinophil count	2.90×10^{-6}	92.24
27	Alkaline phosphatase	8.27×10^{-6}	96.12
28	International standard ratio	2.65×10^{-5}	92.24

[a] ANOVA: analysis of variance.

[b] ADR: available data rate.

Preprocessing

Given the selected 28 biomarkers, the mean number of available biomarkers per patient was 27.22 (SD 2.33) for the training data and 16.86 (SD 1.58) for the testing data, as summarized in Table 3. To handle the missing data, we calculated the mean value from the training data for each biomarker and replaced the missing data with the mean value for the training and testing data sets. We then added two more features (ie, age and gender) to the 28 biomarkers and trained our AI model using 30 features.

Table 3. Number of available blood biomarkers per patient for the artificial intelligence model training.

Data sets and sources	Number of blood biomarkers	
	Mean (SD)	Min-max
Training data: Tongji Hospital	27.22 (2.33)	13-28
Testing data		
Chonnam National University Hospital	20.39 (1.13)	19-24
Wonkwang University Hospital	15.82 (1.94)	14-19
Samsung Medical Center	14.40 (1.58)	14-17
Total	16.86 (1.58)	14-24

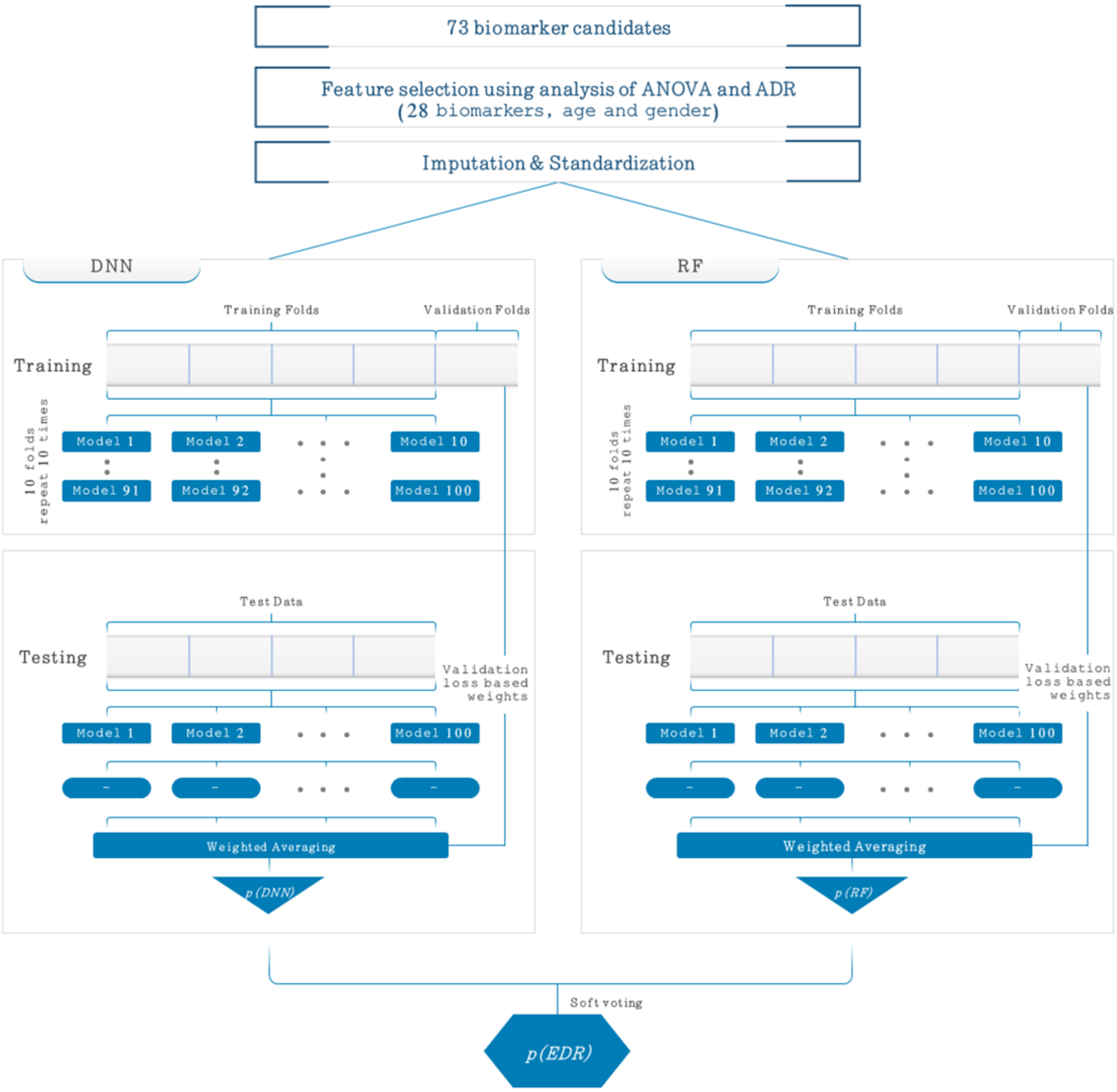
With the 30 features, we performed data set standardization, which is a common requirement for machine learning estimators. The standardization changes the data distribution of each feature with zero mean and standard deviation of 1 as

where *mean(train)* and *SD(train)* are the mean and standard deviation values, respectively, for each feature from the training data. The standardization was applied to the training and testing data sets.

Development of an Ensemble AI Model

As illustrated in Figure 1, the new ensemble AI model is composed of a 5-layer deep neural network (DNN) and RF model. Our ensemble AI model was named as EDRnet (ensemble learning model based on DNN and RF models). The 5-layer DNN was comprised of an input layer, three fully connected (FC) layers, and an output layer. The input layer contained 30 features, including 28 biomarkers, age, and gender. The input layer was fed into three FC layers in a series, each of which consisted of 30, 16, and 8 nodes. To alleviate the overfitting issue, we applied a dropout rate of 0.3. Then, the last FC layer was fed into a softmax layer, which is an output layer providing the probabilities for the patient mortality. Figure S2 in Multimedia Appendix 3 shows our DNN model and its printed textual summary run on Keras, where the total number of parameters (ie, weights and biases) was 1571.

Figure 1. Proposed ensemble model (EDRnet) composed of a 5-layer deep neural network (DNN) and random forest (RF) model for the mortality prediction. In the training of both models, a 10-time-repetition 10-fold stratified cross-validation was separately performed, and the predicted mortality probabilities of the DNN model, p(DNN), and the RF model, p(RF), were calculated. The final predicted mortality probability of the ensemble model, p(EDR), was obtained by soft voting based on the p(DNN) and the p(RF). ADR: available data rate; ANOVA: analysis of variance; EDRnet: ensemble learning model based on DNN and RF models.



For the 5-layer DNN, a 10-time-repetition 10-fold stratified cross-validation was performed to confirm the model's generalization ability. The training data (N=361) were randomly shuffled and partitioned into 10 equal subgroups in a stratified manner. Of the 10 subgroups, a single subgroup was retained as the validation data set for testing the model, and the remaining nine subgroups were used as the training data set. The process was then repeated 10 times, with each of the 10 subgroups used exactly once as the validation data set. By repeating this stratified 10-fold cross-validation process 10 times, a total of 100 models from the 5-layer DNN were derived. Then, we ensembled the models with the weighted average as

where $p_m(DNN)$ is the predicted mortality probability value from the m^{th} model of the DNN, $p(DNN)$ is the ensemble result corresponding to the predicted mortality prediction probability, and ${}_m(DNN)$ is the normalized weight value for the m^{th} model. We obtained the normalized weight value ${}_m(DNN)$ as

where the weight $w_m(DNN)$ was obtained using the validation loss from the m^{th} model, $l_m(DNN)$, as

Along with the 5-layer DNN, we separately trained an RF model. For the RF model, 100 decision trees were trained with a maximum depth of 4 and maximum feature number of 5. Similar to the 5-layer DNN, we performed a 10-time-repetition 10-fold stratified cross-validation and ensembled the 100 models with the weighted average as

where $p_m(RF)$ is the predicted mortality probability value from the m^{th} model of the RF, $p(RF)$ is the ensemble result corresponding to the predicted mortality prediction probability, and $_m(RF)$ is the normalized weight value for the m^{th} model. We obtained the normalized weight value $_m$ as

where the weight $w_m(RF)$ was obtained using the validation loss from the m^{th} model, $l_m(RF)$, as

Given the two ensemble results $p(DNN)$ and $p(RF)$, we finally obtained the final predicted mortality probability value using soft voting. Based on the average of the two probability values $p(DNN)$ and $p(RF)$, if the value is greater than or equal to 0.5, then the prediction result represents death; otherwise, it represents survival.

Implementation

We implemented and trained EDRnet using TensorFlow, version 1.13.1 for graphics processing unit (GPU), and Keras, version 2.2.4 for GPU. NumPy, version 1.16.4; Pandas, version: 0.25.3; Matplotlib, version 3.1.2; and scikit-learn, version 0.22.1, were used to build the model and analyze the results. We trained the models with the Adam optimizer and a binary cross-entropy cost function in equation 9 with a learning rate of 0.0001 and a batch size of 64 on the NVIDIA GeForce GTX 1080 Ti GPU as

where y_i is the label (ie, 1 for deceased and 0 for survived) and $p(y_i)$ is the predicted probability of each patient being deceased for the batch size N number of patients.

Performance Evaluation of AI Models

To evaluate the performance of the AI models in predicting mortality, we used the sensitivity, specificity, accuracy, and balanced accuracy metrics, defined as

where TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative, respectively.

In the training data set, the prediction performance of the AI models was evaluated based on a 10-time-repetition 10-fold stratified cross-validation. In the testing data set, the prediction performance of the AI models was independently evaluated.

To compare the performance of our proposed EDRnet model with those of other external AI models, we separately trained the models of XGBoost and AdaBoost (AB), each of which was evaluated as a single model and as an ensemble model combined with DNN, resulting in four models: XGBoost, AB, ensemble with DNN and XGBoost (EDX), and ensemble with DNN and AB (EDA). For the training of these models, we searched the optimal hyperparameters providing the highest validation accuracy value, as presented in Multimedia Appendix 3, Table S2. In addition, we adopted a recently published AI model by Li et al [6] using a decision tree via an XGBoost-based feature selection for performance comparison. All five external AI models were evaluated using our testing data set of 106 patients.

Results

The cross-validation of RF, DNN, and our ensemble model EDRnet showed that the accuracy on the validation data set is 89% for RF, 92% for DNN, and 93% for EDRnet. Thus, EDRnet provides the highest sensitivity, specificity, accuracy, and balanced accuracy values (see Table 4).

Table 4. Cross-validation accuracy comparison.

Model	Cross-validation results (N=361), mean (SD)			
	Sensitivity	Specificity	Accuracy	Balanced accuracy
Random forest	0.89 (0.06)	0.89 (0.07)	0.89 (0.04)	0.89 (0.04)
Deep neural network	0.91 (0.06)	0.93 (0.04)	0.92 (0.04)	0.92 (0.06)
EDRnet[a]	0.92 (0.05)	0.93 (0.03)	0.93 (0.03)	0.93 (0.05)

[a] EDRnet: ensemble learning model based on deep neural network and random forest models.

Moreover, we applied EDRnet to 106 Korean patients as an independent testing data set to check the TP, TN, FP, FN, sensitivity, specificity, accuracy, and balanced accuracy. The results show a sensitivity of 100%, specificity of 91%, accuracy of 92%, and balanced accuracy of 96%, indicating that the model trained and validated on Chinese patient data can be applied to Korean patients for mortality prediction (see Table 5). The computational times of DNN and RF in EDRnet for the training were 796 and 126 seconds, respectively. The overall computational time for the testing of EDRnet was 72 seconds.

Table 5. Test results from our proposed EDRnet (ensemble learning model based on deep neural network and random forest models) model.

Model	Testing data	True negative, %	False positive, %	False negative, %	True positive, %	Sensitivity	Specificity	Accuracy	Balanced accuracy
EDRnet	CNUH[a]	79	6	0	0	N/A[b]	0.93	0.93	0.93
EDRnet	WKUH[c]	7	2	0	2	1.00	0.78	0.82	0.89
EDRnet	SMC[d]	9	1	0	0	N/A	0.90	0.90	0.90
EDRnet	Total	95	9	0	2	1.00	0.91	0.92	0.96

[a] CNUH: Chonnam National University Hospital.

[b] N/A: not applicable.

[c] WKUH: Wonkwang University Hospital.

[d] SMC: Samsung Medical Center.

Next, we summarized the performance comparison results between XGBoost, AB, RF, DNN, EDX, EDA, Li et al's model [6], and EDRnet. Considering all variables, EDRnet provided the highest prediction performance. Indeed, the balanced accuracy was 88% with XGBoost, 89% with AB, 92% with RF, 71% with DNN, 88% with EDX, 71% with EDA, 67% with Li et al's model [6], and 96% with EDRnet. Notably, the accuracy of Li et al's model was only 36%, indicating that a few blood markers may not be sufficient to predict patient mortality (see Table 6).

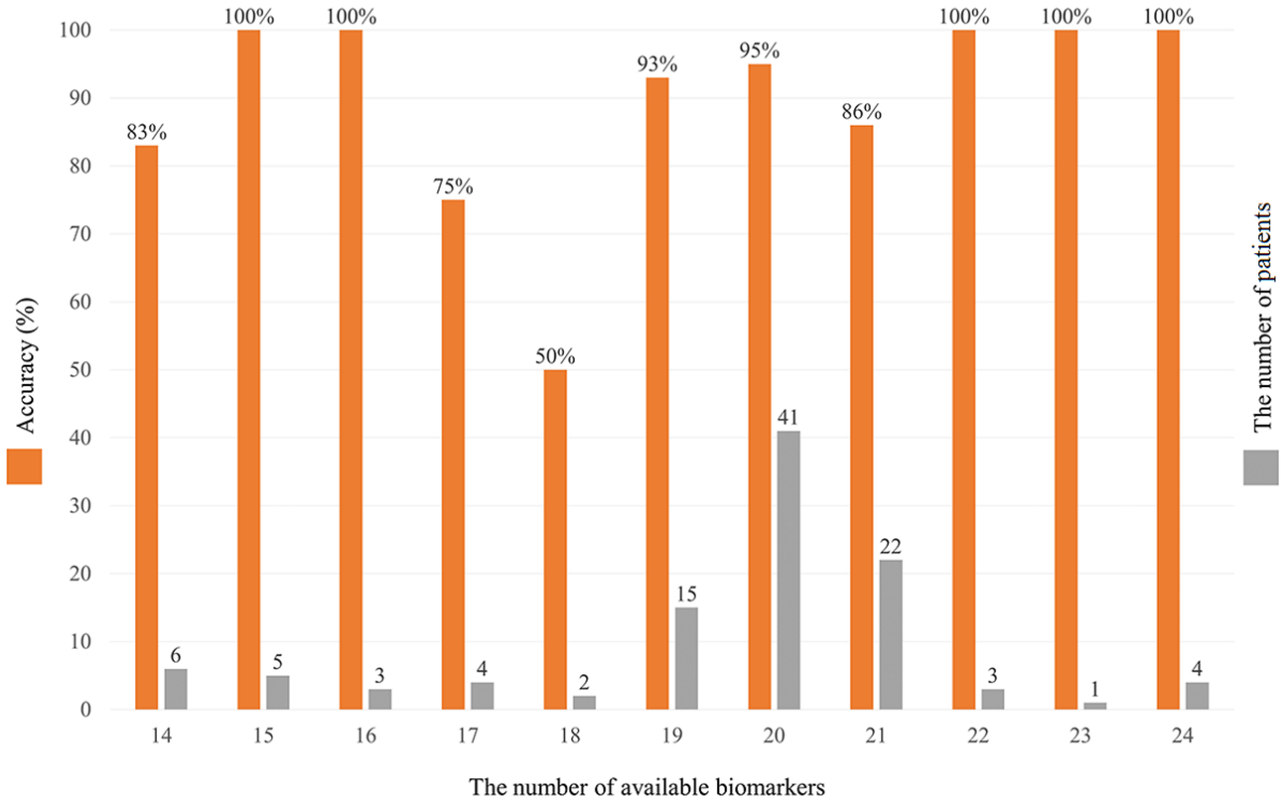
Table 6. Comparison of the performance of various methods.

Model	True negative, %	False positive, %	False negative, %	True positive, %	Sensitivity	Specificity	Accuracy	Balanced accuracy
XGBoost	80	24	0	2	1.00	0.77	0.77	0.88
AdaBoost	81	23	0	2	1.00	0.78	0.78	0.89
Random forest	87	17	0	2	1.00	0.84	0.84	0.92
5-layer deep neural network (DNN)	95	9	1	1	0.50	0.91	0.90	0.71
DNN + XGBoost	80	24	0	2	1.00	0.77	0.77	0.88
DNN + AdaBoost	96	8	1	1	0.50	0.92	0.91	0.71
Li et al's model [6]	36	68	0	2	1.00	0.35	0.36	0.67
DNN + random forest (EDRnet[a])	95	9	0	2	1.00	0.91	0.92	0.96

[a] EDRnet: ensemble learning model based on DNN and random forest models.

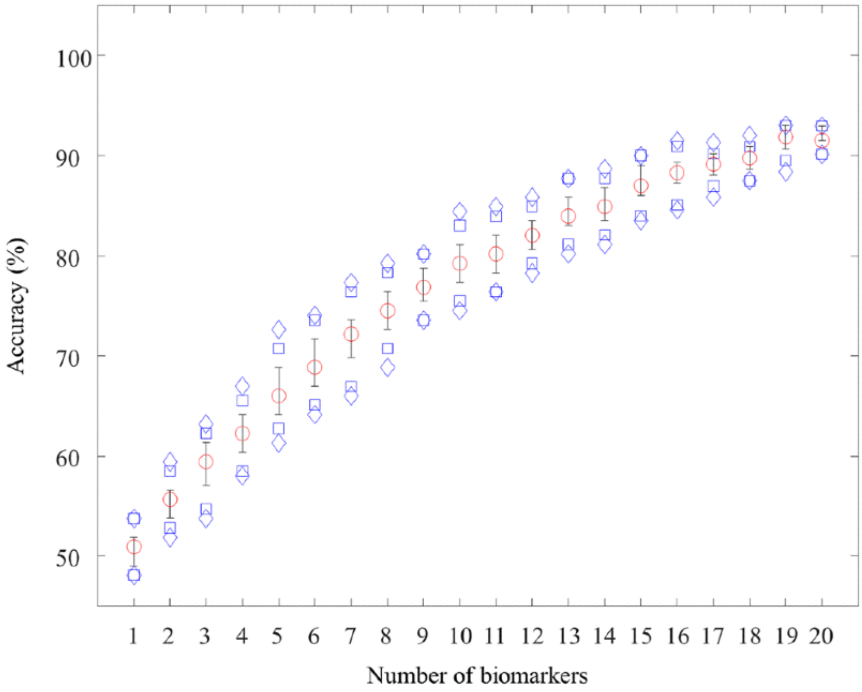
Our proposed EDRnet model used 28 blood biomarkers for prediction, but it does not require all 28 blood biomarkers. In our testing data sets, EDRnet training was validated using available biomarkers, ranging from 14 to 24, for each patient (see Figure 2). The results reveal that the majority of the patients had 19 to 21 available biomarkers (ie, 19 in 15 patients, 20 in 41 patients, and 21 in 22 patients) with a similarly high prediction accuracy (ie, 93%, 95%, and 86%, respectively). For the patients with 17 and 18 available biomarkers, the accuracy was 75% and 50%, respectively. By contrast, the patients with 14 to 16 biomarkers showed a high accuracy ranging from 83% to 100%.

Figure 2. Accuracy with the number of available blood biomarkers from the 106-patient testing data set.



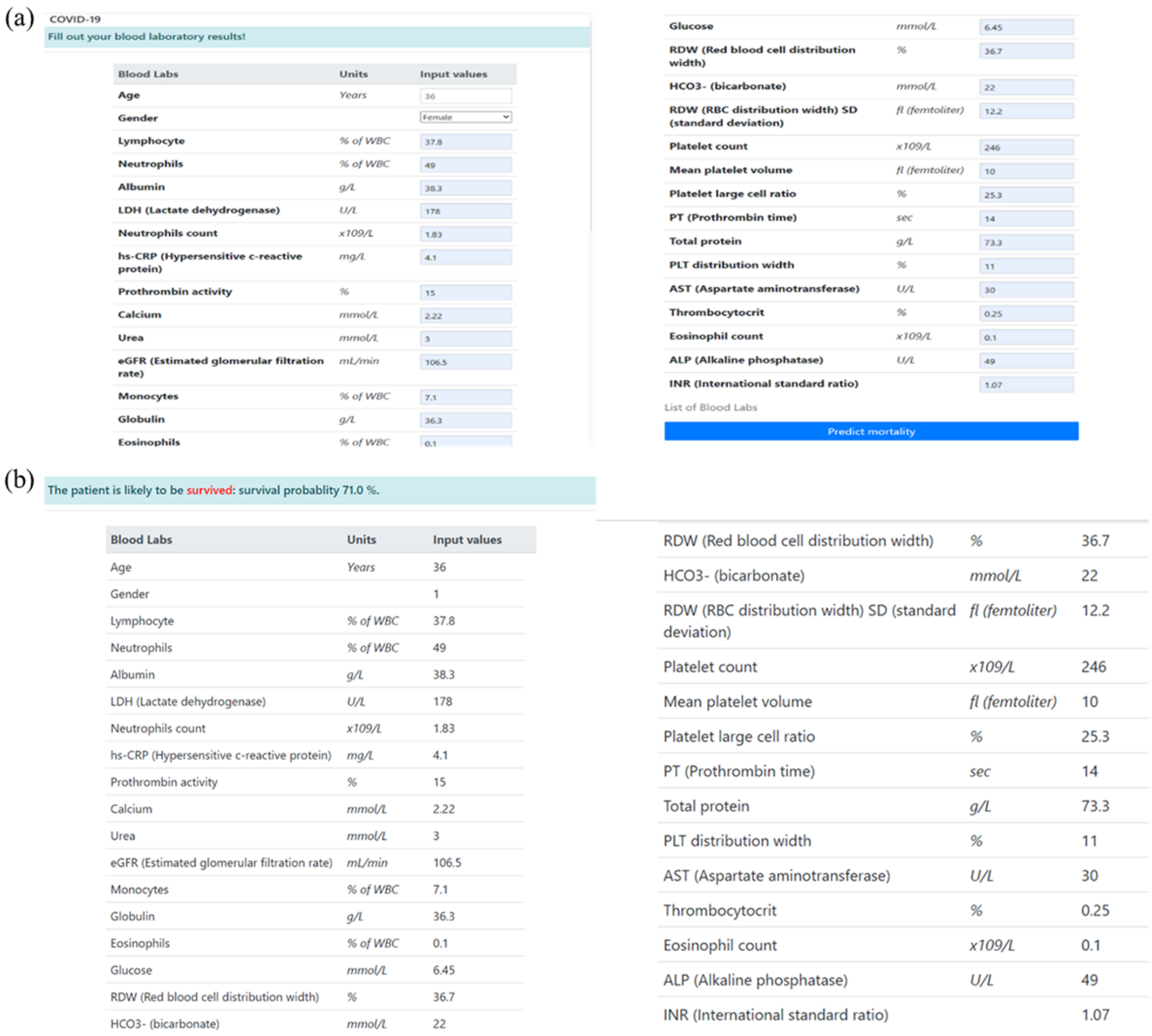
To further investigate the effect of the number of available biomarkers, we estimated the accuracy values according to the number of available biomarkers (see Figure 3). For the estimation, we randomly selected 1 to 20 biomarkers from all of the testing data points and tested the model with a 100-time repetition. When randomly selecting biomarkers, only samples where the actual available number of biomarkers was equal to or greater than the number of randomly selected biomarkers were simulated. The results show that accuracy increases with the number of available biomarkers until reaching 19 biomarkers.

Figure 3. Estimated accuracy values according to the number of available biomarkers. Red circles represent the median. The bars at the top and bottom represent the 75th and 25th percentiles, respectively. The blue rectangles at the top and bottom represent the 90th and 10th percentiles, respectively. The blue diamonds at the top and bottom represent the 95th and 5th percentiles, respectively.



Furthermore, our developed AI model, EDRnet, was successfully deployed on a public website [25] so that anyone can predict mortality using individual blood test results. The web application provides predicted mortality probability, as shown in Figure 4. A user inputs his or her blood sample results (see Figure 4a), and then the predicted mortality results are presented (see Figure 4b). Currently, the web application does not store any information entered by users. However, we consider and plan to store information entered by users on agreement to improve the AI model via a real-time learning process.

Figure 4. Deployed web application, BeatCOVID19 [25]: (a) input windows where a user inputs his or her blood sample results and (b) the predicted mortality results after entering the blood sample results.



(a) COVID-19

Fill out your blood laboratory results!

Blood Labs	Units	Input values
Age	Years	36
Gender		Female
Lymphocyte	% of WBC	37.8
Neutrophils	% of WBC	49
Albumin	g/L	38.3
LDH (Lactate dehydrogenase)	U/L	178
Neutrophils count	x109/L	1.83
hs-CRP (Hypersensitive c-reactive protein)	mg/L	4.1
Prothrombin activity	%	15
Calcium	mmol/L	2.22
Urea	mmol/L	3
eGFR (Estimated glomerular filtration rate)	mL/min	106.5
Monocytes	% of WBC	7.1
Globulin	g/L	36.3
Eosinophils	% of WBC	0.1
Glucose	mmol/L	6.45
RDW (Red blood cell distribution width)	%	36.7
HCO3- (bicarbonate)	mmol/L	22
RDW (RBC distribution width) SD (standard deviation)	fl (femtoliter)	12.2
Platelet count	x109/L	246
Mean platelet volume	fl (femtoliter)	10
Platelet large cell ratio	%	25.3
PT (Prothrombin time)	sec	14
Total protein	g/L	73.3
PLT distribution width	%	11
AST (Aspartate aminotransferase)	U/L	30
Thrombocytocrit	%	0.25
Eosinophil count	x109/L	0.1
ALP (Alkaline phosphatase)	U/L	49
INR (International standard ratio)		1.07

List of Blood Labs

(b) The patient is likely to be survived: survival probablity 71.0 %.

Blood Labs	Units	Input values
Age	Years	36
Gender		1
Lymphocyte	% of WBC	37.8
Neutrophils	% of WBC	49
Albumin	g/L	38.3
LDH (Lactate dehydrogenase)	U/L	178
Neutrophils count	x109/L	1.83
hs-CRP (Hypersensitive c-reactive protein)	mg/L	4.1
Prothrombin activity	%	15
Calcium	mmol/L	2.22
Urea	mmol/L	3
eGFR (Estimated glomerular filtration rate)	mL/min	106.5
Monocytes	% of WBC	7.1
Globulin	g/L	36.3
Eosinophils	% of WBC	0.1
Glucose	mmol/L	6.45
RDW (Red blood cell distribution width)	%	36.7
HCO3- (bicarbonate)	mmol/L	22
RDW (Red blood cell distribution width)	%	36.7
HCO3- (bicarbonate)	mmol/L	22
RDW (RBC distribution width) SD (standard deviation)	fl (femtoliter)	12.2
Platelet count	x109/L	246
Mean platelet volume	fl (femtoliter)	10
Platelet large cell ratio	%	25.3
PT (Prothrombin time)	sec	14
Total protein	g/L	73.3
PLT distribution width	%	11
AST (Aspartate aminotransferase)	U/L	30
Thrombocytocrit	%	0.25
Eosinophil count	x109/L	0.1
ALP (Alkaline phosphatase)	U/L	49
INR (International standard ratio)		1.07

Regarding clinical characteristics (see Table 7), there were no significant differences in comorbidity. In terms of initial symptoms, the deceased group had more frequent dyspnea symptoms than the survivor group (66.7% vs 16.8%; *P*=.04). All patients from the deceased group required oxygen supply. The deceased group had more frequent altered mentality than the survivor group (50.0% vs 1.0%; *P*=.02). There was no significant difference in terms of antiviral drugs (ie, lopinavir or ritonavir, chloroquine or hydroxychloroquine, ribavirin, remdesivir, and oseltamivir) or anti-inflammatory drugs (ie, interferon, dexamethasone, and methylprednisolone) between the deceased and survivor groups. However, the deceased group received more antibiotics or combination therapy.

Table 7. Clinical characteristics of the patient groups from the testing data set.

Characteristics	Deceased group (n=2)	Survivor group (n=104)	Total (N=106)	*P* value
Comorbidity, n (%)				
Diabetes mellitus	0 (0)	10 (9.6)	10 (9.4)	>.99
Asthma	0 (0)	8 (7.7)	8 (7.5)	>.99
Chronic obstructive pulmonary disease	0 (0)	0 (0)	0 (0)	>.99
Coronary heart disease	0 (0)	3 (2.9)	3 (2.8)	>.99
Cardiovascular disease	0 (0)	1 (1.0)	1 (0.9)	>.99
Chronic kidney disease	0 (0)	1 (1.0)	1 (0.9)	>.99
Chronic liver disease	0 (0)	0 (0)	0 (0)	>.99
Congestive heart failure	1 (50)	3 (2.9)	4 (3.8)	.11
Cancer	0 (0)	3 (2.9)	3 (2.8)	>.99
Initial symptom, n (%)				
Fever	2 (100)	59 (56.7)	61 (57.5)	.61
Cough	1 (50)	46 (44.2)	47 (44.3)	>.99
Dyspnea	2 (100)	18 (16.8)	20 (18.2)	.04
Diarrhea	0 (0)	11 (10.3)	11 (10.0)	>.99
Myalgia	0 (0)	26 (24.3)	26 (23.6)	>.99
Initial vital sign, mean (SD)				
Systolic blood pressure (mm Hg)	127.5 (17.7)	128.0 (18.5)	128.0 (18.4)	.97
Diastolic blood pressure (mm Hg)	74.5 (10.6)	78.8 (12.8)	78.7 (12.8)	.64
Heart rate (per minute)	96.5 (24.7)	84.2 (17.7)	84.4 (17.8)	.34
Respiration rate (per minute)	29.5 (10.6)	20.2 (4.1)	20.4 (4.3)	.43
Altered mentality, n (%)	1 (50)	1 (1.0)	2 (1.9)	.02
Oxygen requirement, n (%)				
No oxygen supply	0 (0)	83 (79.8)	83 (78.3)	.07
Conventional oxygen	1 (50)	15 (14.4)	16 (15.1)	.69
High-flow nasal cannula	0 (0)	3 (2.9)	3 (2.8)	>.99
Noninvasive ventilation	0 (0)	0 (0)	0 (0)	>.99
Mechanical ventilation	1 (50)	3 (2.9)	4 (3.8)	.11
Extracorporeal membrane oxygenation	0 (0)	1 (1.0)	1 (0.9)	>.99
Pharmacologic agent, n (%)				
Lopinavir or ritonavir	2 (100)	30 (28.8)	32 (30.2)	.16
Chloroquine or hydroxychloroquine	0 (0)	7 (6.7)	7 (6.6)	>.99
Ribavirin	0 (0)	0 (0)	0 (0)	>.99
Remdesivir	0 (0)	0 (0)	0 (0)	.99
Oseltamivir	0 (0)	2 (1.9)	2 (1.9)	.99
Interferon	0 (0)	0 (0)	0 (0)	>.99
Dexamethasone	0 (0)	1 (1.0)	1 (0.9)	>.99
Methylprednisolone	0 (0)	4 (3.8)	4 (3.8)	>.99
Antibiotics	2 (100)	8 (7.7)	10 (9.4)	.001
Combination	2 (100)	15 (14.4)	17 (16.0)	.02

Discussion

Principal Findings

Our new AI model, EDRnet, was able to predict the mortality of COVID-19 patients using 28 blood biomarkers obtained within 24 hours after hospital admission. In the independent testing data sets, EDRnet showed excellent prediction performance with high sensitivity (100%), specificity (91%), and accuracy (92%). We were able to improve the prediction performance by adopting the ensemble approach combining DNN and RF models. Of note, EDRnet was developed by training with Chinese patients' data and testing with Korean patients' data.

EDRnet has several advantages. First, EDRnet can predict which patients are at a high risk of mortality in the early stage of hospital admission (ie, within 24 hours after admission). This is a substantial improvement compared to the prior AI prediction model reported by Yan et al, which predicted mortality 10 days before the occurrence of survival or death [6]. The mortality prediction at the time of admission can be substantially informative for clinicians because the critical time regarding disease progression is 10 to 14 days from the onset of symptoms, according to previous studies [13,16,26]. EDRnet can provide treatment priority guidance regarding who should be treated intensively. Second, EDRnet only uses blood biomarkers to predict mortality. In general, COVID-19 patients get blood laboratory tests at the time of hospital admission [9,27]. Blood biomarkers are objective indices that are used to estimate patients' conditions in a quantitative manner, which may be beneficial to assure the reliability of the AI model. We did not include subjective biomarkers, such as symptoms, nor predisposing factors, such as underlying comorbidities, because these indices are difficult for quantification and may show high variability between patients. Third, the clinical meaning and significance of blood biomarkers used in our EDRnet model have been well investigated through many prior clinical studies. Thus, the AI's predicted mortality results are explainable and easily understood by doctors. Furthermore, several major blood biomarkers are used in our EDRnet model.

The hematological changes in lymphocytes, neutrophils, monocytes, eosinophils, and platelets are common, as these changes are related to viral replication and hyperinflammation in COVID-19 infection [12,13]. In severe cases, the infiltration and sequestration of CD4+/CD8+ T cells occurred, leading to a decrease in the peripheral lymphocytes. Neutrophil counts [19-21] were significantly higher in the severe group than in the mild group. Platelet count, platelet volume, and platelet large-cell ratio are related to COVID-19 infection because immunologic destruction can lead to inappropriate platelet activation and consumption as well as impaired megakaryopoiesis [28-30].

Regarding blood chemistry, hs-CRP is a major biomarker that represents acute phase inflammation [2,10,14-17]. LDH is related to cell damage, so elevated LDH is an independent risk factor for the severity and mortality of COVID-19 [2,8-11]. Hypoalbuminemia [18,20,21], hypocalcemia [31-33], and elevated aspartate aminotransferase [18] are highly associated with severe COVID-19 infection requiring hospitalization in the intensive care unit. Urea and estimated glomerular filtration rate are important lab findings associated with an underlying chronic renal disease, which is a well-known predisposing factor of mortality [34]. In terms of the coagulation profile, COVID-19 generally presents a hypercoagulation state, thus resulting in an elevated prothrombin time and international normalized ratio in severe COVID-19 cases [3,18].

In this study, no significant differences were observed in the use of pharmacologic agents between the deceased and survivor groups except antibiotics and in the use of antiviral drugs, such as remdesivir. Antibiotics or combination therapy is usually used for suspected bacterial superinfection that represents severe diseases. To date, there has been no successfully effective pharmacologic agent to treat COVID-19. The pharmacologic treatment is not significantly related to survival in this study.

EDRnet does not require all 28 blood biomarkers for the prediction of mortality. EDRnet worked well as long as there were at least 19 blood biomarkers at the time of admission. Compared to prior AI prediction models for COVID-19 mortality, which used three biomarkers, there might be concern that EDRnet requires too many biomarkers. However, these blood tests are commonly performed in our daily clinical practice for hospitalized patients with COVID-19. If more data are accumulated, then we can reduce the number of blood biomarkers for mortality prediction.

Limitations and Future Work

Our study has several limitations. First, the number of patients available for testing might be small. According to Johns Hopkins Coronavirus Resource Center, the mortality rate in South Korea is 1.7%. In the testing data set of 106 Korean patients, the mortality rate was 1.9%, which is almost equivalent to the actual mortality rate. It might be necessary to update EDRnet by training with a large population data set from all over the world. To update EDRnet, we made a web application [25] so that anyone can access the model. We believe that opening the AI model to the public is helpful to improve its performance and generalizability. Second, our data did not include other races, such as Caucasian or Middle East Asian. Our future research plan is to establish a real-time AI training system that can continue to train our model using prospectively collected data from all over the world. In addition, we will upgrade the web application so that the database framework allows a user to input his or her blood sample results along with the outcome. Based on the extended data, we will improve EDRnet for better generalization.

Conclusions

In conclusion, our new AI model, EDRnet, was developed to predict the mortality of COVID-19 patients at the time of hospital admission using blood biomarkers only. It is now open to the public with the hope that it can help health care providers fight COVID-19 and improve patients' outcomes.

Acknowledgments

This work was supported by Korea Medical Device Development Fund grants, funded by the Korean government (the Ministry of Science and ICT; the Ministry of Trade, Industry and Energy; the Ministry of Health and Welfare, Republic of Korea; and the Ministry of Food and Drug Safety) (grant Nos. 202012B04 and NRF-2020R1A2C1014829), and by the Korea Health Industry Development Institute (grant No. HI18C1216). Author RJ is currently an MIT Lincoln Laboratory employee. No Laboratory funding or resources were used to produce the result/findings reported in this publication.

Authors' Contributions

HK and HC carried out the machine learning and deep learning simulation for hyperparameter search and modeling. CP verified blood biomarkers to be applied in COVID-19 patients and wrote an initial draft of the paper. DWK, SEK, CRC, and REK collected and validated the data and performed statistical analyses. HL developed and maintained the web application. KWK, JHS, and TYC validated and confirmed the simulations and helped to draft the manuscript. RJ validated the methodology of machine learning and deep learning techniques. JL and WSK conceived of the study and participated in its design and coordination and wrote the initial manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Training data set.
[XLSX File (Microsoft Excel File), 96 KB - jmir_v22i12e25442_app1.xlsx]

Multimedia Appendix 2
Testing data set.
[XLSX File (Microsoft Excel File), 35 KB - jmir_v22i12e25442_app2.xlsx]

Multimedia Appendix 3
Supplementary figures and tables.
[DOCX File , 876 KB - jmir_v22i12e25442_app3.docx]

References

1. COVID-19 Surveillance Group. Characteristics of COVID-19 Patients Dying in Italy: Report Based on Available Data on March 20th, 2020. Rome, Italy: Istituto Superiore di Sanità; 2020. URL: https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_20_marzo_eng.pdf [accessed 2020-12-17]
2. Li C, Ye J, Chen Q, Hu W, Wang L, Fan Y, et al. Elevated lactate dehydrogenase (LDH) level as an independent risk factor for the severity and mortality of COVID-19. Aging (Albany NY) 2020 Aug 14;12(15):15670-15681 [FREE Full text] [doi: 10.18632/aging.103770] [Medline: 32805722]
3. Luo L, Xu M, Du M, Kou H, Liao D, Cheng Z, et al. Early coagulation tests predict risk stratification and prognosis of COVID-19. Aging (Albany NY) 2020 Aug 29;12(16):15918-15937 [FREE Full text] [doi: 10.18632/aging.103581] [Medline: 32860672]
4. Lai C, Liu YH, Wang C, Wang Y, Hsueh S, Yen M, et al. Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Facts and myths. J Microbiol Immunol Infect 2020 Jun;53(3):404-412 [FREE Full text] [doi: 10.1016/j.jmii.2020.02.012] [Medline: 32173241]
5. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
6. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2020 May 14;2(5):283-288. [doi: 10.1038/s42256-020-0180-7]
7. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: Association for Computing Machinery; 2016 Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA p. 785-794 URL: https://dl.acm.org/doi/pdf/10.1145/2939672.2939785 [doi: 10.1145/2939672.2939785]
8. Wu MY, Yao L, Wang Y, Zhu XY, Wang XF, Tang PJ, et al. Clinical evaluation of potential usefulness of serum lactate dehydrogenase (LDH) in 2019 novel coronavirus (COVID-19) pneumonia. Respir Res 2020 Jul 06;21(1):171 [FREE Full text] [doi: 10.1186/s12931-020-01427-8] [Medline: 32631317]

9. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): A meta-analysis. Clin Chem Lab Med 2020 Jun 25;58(7):1021-1028. [doi: 10.1515/cclm-2020-0369] [Medline: 32286245]
10. Henry BM, Aggarwal G, Wong J, Benoit S, Vikse J, Plebani M, et al. Lactate dehydrogenase levels predict coronavirus disease 2019 (COVID-19) severity and mortality: A pooled analysis. Am J Emerg Med 2020 Sep;38(9):1722-1726 [FREE Full text] [doi: 10.1016/j.ajem.2020.05.073] [Medline: 32738466]
11. Panteghini M. Lactate dehydrogenase: An old enzyme reborn as a COVID-19 marker (and not only). Clin Chem Lab Med 2020 Aug 24;58(12):1979-1981. [doi: 10.1515/cclm-2020-1062] [Medline: 32829312]
12. Huang I, Pranata R. Lymphopenia in severe coronavirus disease-2019 (COVID-19): Systematic review and meta-analysis. J Intensive Care 2020;8:36 [FREE Full text] [doi: 10.1186/s40560-020-00453-4] [Medline: 32483488]
13. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med 2020 May;46(5):846-848 [FREE Full text] [doi: 10.1007/s00134-020-05991-x] [Medline: 32125452]
14. Wang L. C-reactive protein levels in the early stage of COVID-19. Med Mal Infect 2020 Jun;50(4):332-334 [FREE Full text] [doi: 10.1016/j.medmal.2020.03.007] [Medline: 32243911]
15. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
16. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
17. Tan C, Huang Y, Shi F, Tan K, Ma Q, Chen Y, et al. C-reactive protein correlates with computed tomographic findings and predicts severe COVID-19 early. J Med Virol 2020 Jul;92(7):856-862 [FREE Full text] [doi: 10.1002/jmv.25871] [Medline: 32281668]
18. Long H, Nie L, Xiang X, Li H, Zhang X, Fu X, et al. D-dimer and prothrombin time are the significant indicators of severe COVID-19 and poor prognosis. Biomed Res Int 2020;2020:6159720 [FREE Full text] [doi: 10.1155/2020/6159720] [Medline: 32596339]
19. Kong M, Zhang H, Cao X, Mao X, Lu Z. Higher level of neutrophil-to-lymphocyte is associated with severe COVID-19. Epidemiol Infect 2020 Jul 09;148:e139 [FREE Full text] [doi: 10.1017/S0950268820001557] [Medline: 32641174]
20. Aziz M, Fatima R, Lee-Smith W, Assaly R. The association of low serum albumin level with severe COVID-19: A systematic review and meta-analysis. Crit Care 2020 May 26;24(1):255 [FREE Full text] [doi: 10.1186/s13054-020-02995-3] [Medline: 32456658]
21. Aziz M, Fatima R, Assaly R. Elevated interleukin-6 and severe COVID-19: A meta-analysis. J Med Virol 2020 Nov;92(11):2283-2285 [FREE Full text] [doi: 10.1002/jmv.25948] [Medline: 32343429]
22. The Lancet Haematology. COVID-19 coagulopathy: An evolving story. Lancet Haematol 2020 Jun;7(6):e425 [FREE Full text] [doi: 10.1016/S2352-3026(20)30151-4] [Medline: 32470428]
23. Arachchillage DRJ, Laffan M. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost 2020 May;18(5):1233-1234. [doi: 10.1111/jth.14820] [Medline: 32291954]
24. Bowles L, Platton S, Yartey N, Dave M, Lee K, Hart DP, et al. Lupus anticoagulant and abnormal coagulation tests in patients with COVID-19. N Engl J Med 2020 Jul 16;383(3):288-290 [FREE Full text] [doi: 10.1056/NEJMc2013656] [Medline: 32369280]
25. BeatCOVID19. URL: http://beatcovid19.ml/ [accessed 2020-12-15]
26. Qian G, Yang N, Ding F, Ma AHY, Wang Z, Shen Y, et al. Epidemiologic and clinical characteristics of 91 hospitalized patients with COVID-19 in Zhejiang, China: A retrospective, multi-centre case series. QJM 2020 Jul 01;113(7):474-481 [FREE Full text] [doi: 10.1093/qjmed/hcaa089] [Medline: 32181807]
27. Luo Y, Yuan X, Xue Y, Mao L, Lin Q, Tang G, et al. Using a diagnostic model based on routine laboratory tests to distinguish patients infected with SARS-CoV-2 from those infected with influenza virus. Int J Infect Dis 2020 Jun;95:436-440 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.078] [Medline: 32371192]
28. Zhao X, Wang K, Zuo P, Liu Y, Zhang M, Xie S, et al. Early decrease in blood platelet count is associated with poor prognosis in COVID-19 patients-Indications for predictive, preventive, and personalized medical approach. EPMA J 2020 May 14:1-7 [FREE Full text] [doi: 10.1007/s13167-020-00208-z] [Medline: 32419876]
29. Zhu S, Dong L, Cai W. Predictive value of neutrophil to lymphocyte and platelet to lymphocyte ratio in COVID-19. Crit Care 2020 Aug 28;24(1):532 [FREE Full text] [doi: 10.1186/s13054-020-03258-x] [Medline: 32859254]
30. Sadr S, SeyedAlinaghi S, Ghiasvand F, Hassan Nezhad M, Javadian N, Hossienzade R, et al. Isolated severe thrombocytopenia in a patient with COVID-19: A case report. IDCases 2020;21:e00820 [FREE Full text] [doi: 10.1016/j.idcr.2020.e00820] [Medline: 32483524]
31. Cappellini F, Brivio R, Casati M, Cavallero A, Contro E, Brambilla P. Low levels of total and ionized calcium in blood of COVID-19 patients. Clin Chem Lab Med 2020 Aug 27;58(9):e171-e173. [doi: 10.1515/cclm-2020-0611] [Medline: 32459190]

32. Di Filippo L, Formenti AM, Rovere-Querini P, Carlucci M, Conte C, Ciceri F, et al. Hypocalcemia is highly prevalent and predicts hospitalization in patients with COVID-19. Endocrine 2020 Jun;68(3):475-478 [FREE Full text] [doi: 10.1007/s12020-020-02383-5] [Medline: 32533508]
33. Sun J, Zhang W, Zou L, Liu Y, Li J, Kan X, et al. Serum calcium as a biomarker of clinical severity and prognosis in patients with coronavirus disease 2019. Aging (Albany NY) 2020 Jun 25;12(12):11287-11295 [FREE Full text] [doi: 10.18632/aging.103526] [Medline: 32589164]
34. Gao M, Wang Q, Wei J, Zhu Z, Li H. Severe coronavirus disease 2019 pneumonia patients showed signs of aggravated renal impairment. J Clin Lab Anal 2020 Oct;34(10):e23535 [FREE Full text] [doi: 10.1002/jcla.23535] [Medline: 32840917]

Abbreviations

AB: AdaBoost
ADR: available data rate
AI: artificial intelligence
ANOVA: analysis of variance
CNUH: Chonnam National University Hospital
DNN: deep neural network
EDA: ensemble with deep neural network and AdaBoost
EDRnet: ensemble learning model based on deep neural network and random forest models
EDX: ensemble with deep neural network and XGBoost
FC: fully connected
FN: false negative
FP: false positive
GPU: graphics processing unit
hs-CRP: high-sensitivity C-reactive protein
LDH: lactate dehydrogenase
RF: random forest
SMC: Samsung Medical Center
TN: true negative
TP: true positive
WKUH: Wonkwang University Hospital

Edited by G Eysenbach, R Kukafka; submitted 03.11.20; peer-reviewed by J Lee, C Jeong; comments to author 24.11.20; revised version received 24.11.20; accepted 08.12.20; published 23.12.20.

Please cite as:
Ko H, Chung H, Kang WS, Park C, Kim DW, Kim SE, Chung CR, Ko RE, Lee H, Seo JH, Choi TY, Jaimes R, Kim KW, Lee J
An Artificial Intelligence Model to Predict the Mortality of COVID-19 Patients at Hospital Admission Time Using Routine Blood Samples: Development and Validation of an Ensemble Model
J Med Internet Res 2020;22(12):e25442
URL: http://www.jmir.org/2020/12/e25442/
doi:10.2196/25442
PMID:33301414

Original Paper

Improving Provision of Preanesthetic Information Through Use of the Digital Conversational Agent “MyAnesth”: Prospective Observational Trial

Fabrice Ferré[1], MD, MSc; Nicolas Boeschlin[1], MD; Bruno Bastiani[2], PhD; Adeline Castel[1], MD; Anne Ferrier[1], MD; Laetitia Bosch[1], MD; Fabrice Muscari[3], MD, PhD; Matt Kurrek[1,4], MD; Olivier Fourcade[1], MD, PhD; Antoine Piau[5], MD, PhD; Vincent Minville[1], MD, PhD

[1]Département d’Anesthésie-Réanimation, Hopital Pierre-Paul Riquet, CHU Purpan, Toulouse, France
[2]Unité Mixte de Recherche Éducation, Formation, Travail, Savoirs (UMR EFTS), Université Toulouse Jean Jaurès, Toulouse, France
[3]Département de chirurgie digestive et transplantation d’organes, CHU Rangueil, Toulouse, France
[4]Department of Anesthesia, University of Toronto, Toronto, ON, Canada
[5]Département de Gériatrie, CHU Rangueil, Toulouse, France

Corresponding Author:
Fabrice Ferré, MD, MSc
Département d’Anesthésie-Réanimation
Hopital Pierre-Paul Riquet
CHU Purpan
Place du Dr Baylac
Toulouse, 31059
France
Phone: 33 0667586887
Email: fabriceferre31@gmail.com

Abstract

Background: Due to time limitations, the preanesthetic consultation (PAC) is not the best time for patients to integrate information specific to their perioperative care pathway.

Objective: The main objectives of this study were to evaluate the effectiveness of a digital companion on patients' knowledge of anesthesia and their satisfaction after real-life implementation.

Methods: We conducted a prospective, monocentric, comparative study using a before-and-after design. In phase 1, a 9-item self-reported anesthesia knowledge test (Delphi method) was administered to patients before and after their PAC (control group: PAC group). In phase 2, the study was repeated immediately after the implementation of a digital conversational agent, MyAnesth (@+PAC group). Patients’ satisfaction and their representations for anesthesia were also assessed using a Likert scale and the Abric method of hierarchized evocation.

Results: A total of 600 tests were distributed; 205 patients and 98 patients were included in the PAC group and @+PAC group, respectively. Demographic characteristics and mean scores on the 9-point preinformation test (PAC group: 4.2 points, 95% CI 3.9-4.4; @+PAC: 4.3 points, 95% CI 4-4.7; *P*=.37) were similar in the two groups. The mean score after receiving information was better in the @+PAC group than in the PAC group (6.1 points, 95% CI 5.8-6.4 points versus 5.2 points, 95% CI 5.0-5.4 points, respectively; *P*<.001), with an added value of 0.7 points (95% CI 0.3-1.1; *P*<.001). Among the respondents in the @+PAC group, 82% found the information to be clear and appropriate, and 74% found it easily accessible. Before receiving information, the central core of patients’ representations for anesthesia was focused on the fear of being put to sleep and thereafter on caregiver skills and comfort.

Conclusions: The implementation of our digital conversational agent in addition to the PAC improved patients' knowledge about their perioperative care pathway. This innovative audiovisual support seemed clear, adapted, easily accessible, and reassuring. Future studies should focus on adapting both the content and delivery of a digital conversational agent for the PAC in order to maximize its benefit to patients.

(J Med Internet Res 2020;22(12):e20455) doi:10.2196/20455

KEYWORDS

chatbot; digital conversational agent; preanesthetic consultation; Abric method; eHealth; digital health; anesthesia

Introduction

Currently, in France, a patient requiring scheduled surgery must go through several mandatory steps: a consultation with the surgeon, a consultation with the nurse, a preanesthetic consultation (PAC), and a preanesthetic history and physical examination.

The PAC became compulsory in France on December 5, 1994, by Decree No. 94-1050, which stated that the consultation must be “led by an anesthetist physician” who sets out an anesthesia protocol [1]. In 2018, anesthetists at University Hospital Center of Toulouse (Purpan Hospital, Toulouse, France) performed nearly 12,000 PACs, including 4500 for elective orthopedic surgery (representing over 1500 hours per year devoted to PACs in this unit).

Because only 15 to 20 minutes can be devoted to the PAC per patient, only a few minutes are dedicated to the explanation of the anesthetic (its advantages and disadvantages, and risks and alternatives, if they exist), as well as of the risks inherent in their conditions and the possible ways to reduce them [2]. The amount of new information the patient must process appears disproportionately large when compared with the short duration of the consultation. Moreover, the context of a consultation is a source of anxiety (eg, “white coat effect”) and can thus prove deleterious to the retention of such information.

At the same time, the multiplicity of tasks incumbent on anesthetists reduces the time available, which may explain why the time devoted to presenting patients with information during PACs is often reduced [3]. Providing information tailored to each patient and each situation and ensuring that it is well understood is a daily challenge. In addition, the patient’s knowledge about the anesthesia often appears limited. For instance, a 1994 study by Swinhoe and Groves [4] showed that 35% of the patients did not know that the anesthetist was a physician.

Recently, digital conversational agents (also known as chatbots) have been emerging in the health care field, including in the management of complex older populations [5]. These digital companions are very useful for communicating with the patient before or after care without overloading the clinicians. They have the benefit of being available at any time and can be used repeatedly, at home or elsewhere. In this setting, Bibault et al [6] were able to demonstrate that the quality of breast cancer information delivered by a digital conversational agent was equivalent to a specialized consultation. The interest in this type of approach as compared with the unsupervised open access to information on the internet is the ability to control the content and the accuracy of the information offered.

The main objective of this study was to develop a digital companion that could help patients to prepare for their scheduled orthopedic surgery by providing them with adapted information before their PAC.

We hypothesized that the implementation of this tool, before and in addition to the PAC, would improve the quality of the information delivered in comparison with the standard practice.

Methods

Experimental Design

In this before-and-after study, we planned two successive phases that allowed us to define two groups.

The first phase took place before the implementation of the digital conversational agent. Patients were evaluated by the test before and immediately after the PAC. This control group was referred to as the PAC group.

During the second phase of the study, access to the digital conversational agent, or chatbot, was granted at the moment of the surgical decision and until the PAC. The evaluation by the test was conducted before access to the chatbot and immediately after the PAC. This intervention group was referred to as the @+PAC group.

Timeline of the Study

Phase 1 (PAC group) was carried out for 3 months, from February 1, 2019, to April 30, 2019. Phase 2 (@+PAC group) was carried out for 3 months, from June 1, 2019, to August 31, 2019. The anesthetists in the PACs were not informed of this timeline.

Population

We included patients aged 18 to 85 years who were scheduled for a PAC before elective orthopedic surgery at the University Hospital Center of Toulouse (Purpan Hospital, Toulouse, France). The patient exclusion criteria were (1) having the PAC in a different hospital, (2) the presence of a major sensory handicap (blindness or deafness) compromising the comprehension of the information, or (3) the inability to give informed consent.

During the PAC, the physician consulted the medical documents brought by the patient, questioned and examined the patient, and informed the patient of the benefits and associated risks of the anesthetic procedures. The anesthetist could ask for complementary investigations if necessary. In addition, an information booklet about anesthetic techniques was given to the patient.

Demographic Data Collection

Age, sex, height and weight, profession, education (number of postgraduate years completed), number of previous anesthetics received, smartphone use, type of surgery, and type of hospitalization (outpatient or inpatient) were anonymously collected.

Interventions

The digital conversational agent MyAnesth was developed in collaboration with a company creating secure health companions (BOTdesign, Toulouse, France).

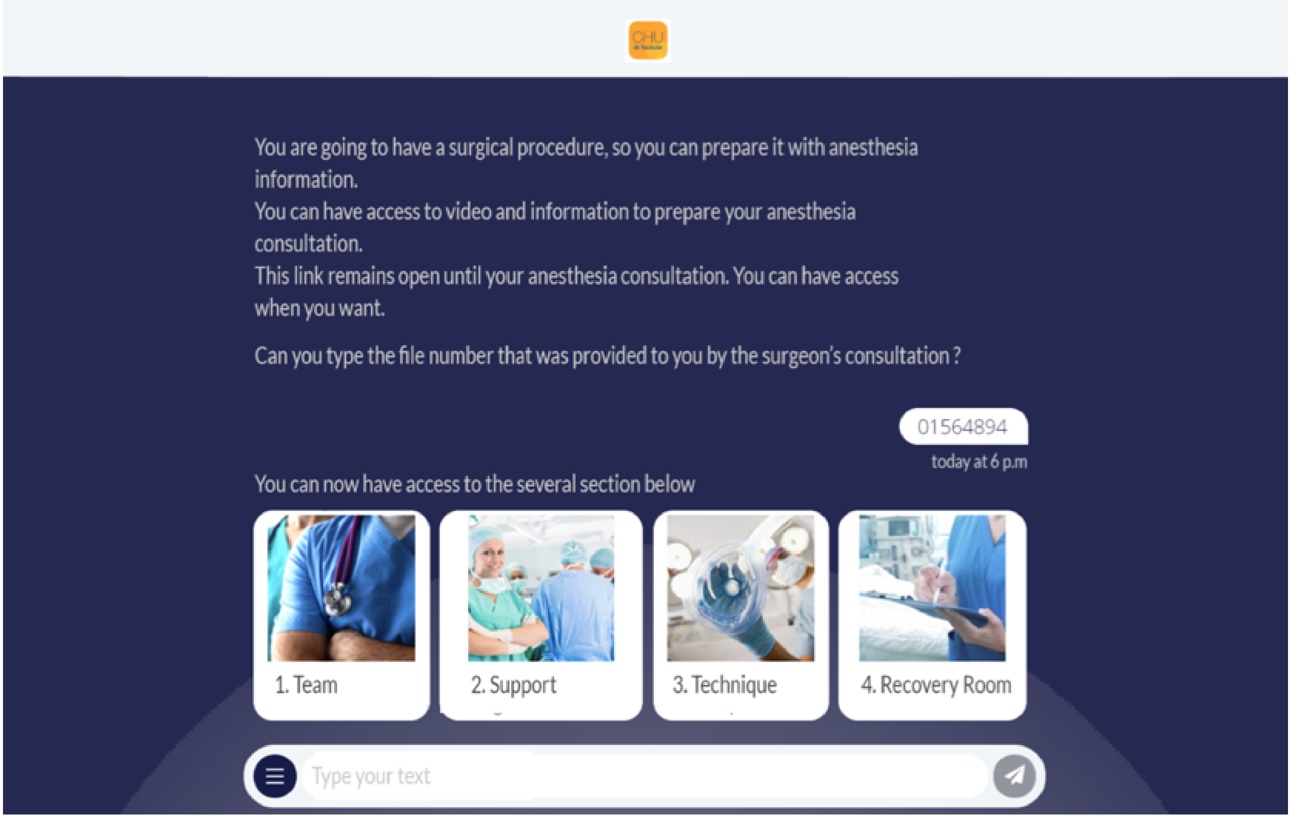
Its content was developed by 6 anesthetists from the University Hospital Center of Toulouse (orthopedic surgery unit, Purpan Hospital, Toulouse, France), taking into account data from the literature on fears generated by anesthesia [7-9]. Information considered important to be delivered to patients was then the subject of a team consensus.

The wording of the informative messages, such as frequently asked questions (FAQs) and their answers, were revised by a specialist in social and human sciences at Paul Sabatier University in Toulouse, France.

Access to the conversational agent was made possible by a URL link sent to the patient and accessible from any electronic device (ie, smartphone, tablet, or computer). This digital tool complies with all French and European regulations in terms of health data security.

Each patient could browse through 4 themed sections in the order of their choice: (1) team, (2) support, (3) technique, and (4) recovery room (Figure 1). Each section included a video, and the most FAQs and their answers, which could be accompanied by a picture (Multimedia Appendix 1). Within the support section, the content was adapted to the type of hospitalization (outpatient or inpatient), which provided the patient with more personalized information. A synthesized voice read the written information (eg, terms and conditions of use, and FAQs). The videos were subtitled in French to optimize patients' comprehension.

Figure 1. Access page to the digital conversational agent MyAnesth (BOTdesign, Toulouse, France).



You are going to have a surgical procedure, so you can prepare it with anesthesia information.
You can have access to video and information to prepare your anesthesia consultation.
This link remains open until your anesthesia consultation. You can have access when you want.

Can you type the file number that was provided to you by the surgeon's consultation ?

01564894
today at 6 p.m

You can now have access to the several section below

A version of the digital conversational agent, for demonstration purposes only, and identical to the one used in the study, is accessible online [10].

Primary Outcome: Self-Reported Anesthesia Knowledge Test

The information was considered successfully delivered if there was a significant increase in the score on the self-reported anesthesia knowledge test (Figure 2). This test consisted of 9 multiple-choice questions developed using the Delphi method [11]. The questions were simple and considered by the panel of experts as constituting the minimum amount of knowledge required before anesthesia. Each multiple-choice question had 4 answers, including the option "I don't know." Only one correct answer (1 point) was possible for each multiple-choice question. A wrong answer or "I don't know" response was worth 0 points, and the total score was between 0 and 9 points.

Figure 2. Self-reported anesthesia knowledge test. The correct answers are 1A, 2B, 3A, 4B, 5B, 6C, 7B, 8A, and 9C.

1. **I can drink:**
A. Up to 2 hours before the intervention
B. Unrestricted until operation
C. Until the evening before
D. I don't know

2. **If the anesthesiologist performs a regional anesthesia:**
A. I will not have pain medication
B. I will need less pain medication
C. Pain medication will be prohibited
D. I don't know

3. **In the event of a complication:**
A. I will be informed
B. My length of hospital stay cannot be changed
C. There is never a complication
D. I don't know

4. **Regional anesthesia:**
A. Just lasts the time of the operation
B. Lasts longer than the operation
C. Lasts less than the operation
D. I don't know

5. **During general anesthesia:**
A. I breathe as usual
B. I breathe through a hose connected to a mechanical respirator
C. I am not breathing
D. I don't know

6. **After the intervention:**
A. My attention may be diminished during the first hours
B. My attention will not be affected by anesthesia
C. My attention can be reduced for several days
D. I don't know

7. **Tobacco:**
A. Has no consequences with anesthesia
B. Should be stopped at least four weeks before anesthesia
C. Should be stopped the day before anesthesia
D. I don't know

8. **Regional anesthesia:**
A. Is performed using an ultrasound scan
B. Has no risk
C. Is performed by the nurse anesthetist
D. I don't know

9. **The day of surgery:**
A. I will see the same anesthesiologist as the one in consultation
B. I will not see an anesthesiologist
C. I will see one of the members of the anesthesiology medical team
D. I don't know

Secondary Outcomes

Patients' Satisfaction

To evaluate patients' satisfaction regarding the acceptability and quality of the delivered information, patients were required to answer questions rated from 1 to 5 according to the Likert method (1=strongly disagree, 2=rather disagree, 3=no opinion, 4=rather agree, and 5=strongly agree) (Multimedia Appendix 2).

Patients' Representation

In order to assess the patient's representational field for anesthesia, we used the Abric method of hierarchized evocation [12]. This method consists of asking patients a question that stimulates them to respond with 3 words or expressions associated with an inductor (word, sentence, or idea). In our case, the inductor was the word anesthesia, which was introduced by the following question: "What are the 3 words that come to your mind when you hear about anesthesia?" Patients were asked to answer this question before and after receiving information. The written order of the patient's answers expressed the level of importance of each response (first row being the most important, third row the least important).

This method allowed us to come closer to the subject's representations by dividing them into the core of the representation and the peripheral elements. The core constitutes "a mental filter through which reality is perceived and judged" [12]. It is rather independent from the context, contrary to the peripheral elements, which adapt the core to the diversity of the context [12].

The treatment of these data consisted of an analysis that took into account the frequencies of response of a word and of the written order of the responses [13,14]. This analysis allowed us to identify the core zone and the first periphery corresponding to the strongest frequencies in rows 1 and 2, followed by the contrasted elements zone and the second periphery corresponding to the weakest frequencies and row 3 [14].

Sample Size Projection

No data were available from the literature to allow us to calculate a priori the sample size required to identify an increase in the anesthesia knowledge test score. In a pilot study conducted on 30 patients, we identified an increase of 1.0 point (SD 1.3 points) in test scores taken before and after the PAC (4 of 9 points and 5 of 9 points, respectively) on the anesthesia knowledge test.

Considering that the implementation of the digital companion could allow the gain of 1 more point (increase judged to be clinically relevant) compared with the PAC alone, we calculated that 48 patients per phase would be required to demonstrate this difference with a type I error of 5% and a power of 90%. Taking into account the number of patients potentially lost, the technical difficulties inherent to the use of a digital conversational agent not yet tested, and the total number of patients seen in our PACs, we planned for 2 successive periods of 3 months each to include all eligible patients.

Statistical Analysis

The normality of the data was assessed using the Shapiro-Wilk test. The qualitative data were expressed as numbers (percentages). The quantitative data were expressed as median (range) or mean (SD) as appropriate. The categorical variables were compared using the Fisher exact test or the chi-square test. Continuous variables were compared using the Wilcoxon signed-rank test or the Student t test as appropriate. The statistical analysis was done using MedCalc Statistical Software, version 12.6.1 (MedCalc Software bvba, Ostend, Belgium). $P<.05$ was considered statistically significant.

Ethics

The connection to the digital conversational agent was made anonymous by a 4-digit number delivered by one of the investigating physicians during the interview with the programming nurse. Even though no information about the participants' health condition was asked at the time of the connection, the company BOTdesign (Toulouse, France) had access neither to the patients' identity nor to their internet protocol address. This strategy of data protection was decided in agreement with the eHealth committee of the University Center Hospital of Toulouse.

This research was considered to be an experimentation in educational sciences looking to (1) evaluate the quality of the information delivered through an innovative pedagogical tool, and (2) investigate the participants' satisfaction. Hence, this research was deemed to fall outside the Jardé law. For each patient, one of the investigating physicians delivered information about the methods of this research and ensured their nonopposition to participate. The lack of return of the questionnaire was considered a refusal to participate.

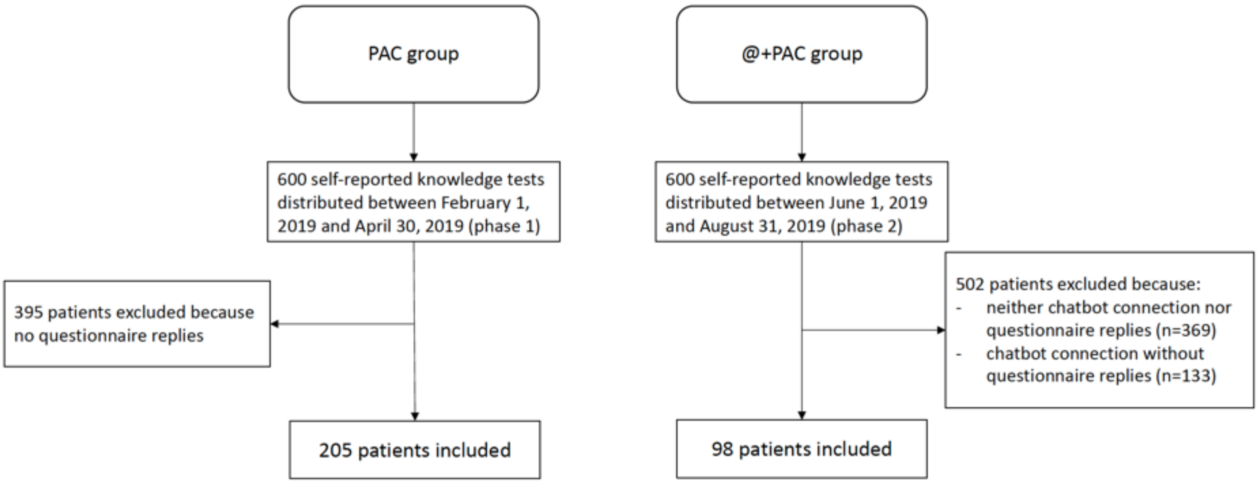
This study did not present any risk to the participants, nor did it modify the regular care process or the time require to care for the patients.

Results

Patients' Characteristics

A total of 303 patients completed the questionnaire and were analyzed during the study period. Of these, 205 patients were included in the PAC group (phase 1) and 98 were included in the @+PAC group (phase 2). The flow chart of patient selection is presented in Figure 3.

Figure 3. Flow chart of patient selection. PAC: preanesthetic consultation; @+PAC: digital conversational agent and preanesthetic consultation.



Demographic characteristics are presented in Table 1. The number of previous anesthetics received and the number of years of postgraduate education were comparable between the groups. Most patients included in the study were outpatients. Both groups had a high rate of internet access on their smartphones or at home (Table 1).

Table 1. Patients' characteristics.

	PAC[a] group (n=205)	@+PAC[b] group (n=98)	*P* value
Age (years), mean (range)	48 (18-85)	50 (18-74)	.73
Female sex, n (%)	95 (46.3)	41 (41.8)	.54
Number of previous anesthetics received, mean (SD)	4 (3.7)	3.9 (2.8)	.79
Number of postgraduate years of education, mean (SD)	1.4 (1.9)	1.5 (1.9)	.70
Internet access at home, n (%)	187 (91.2)	91 (92.9)	.64
Internet access with smartphone, n (%)	174 (84.9)	86 (87.8)	.60
Outpatient, n (%)	137 (66.8)	71 (72.4)	.22

[a] PAC: preanesthetic consultation.

[b] @+PAC: digital conversational agent and preanesthetic consultation.

Anesthesia Knowledge Test

The results of the anesthesia knowledge test are shown in Table 2 and illustrated in Figure 4. The implementation of the digital conversational agent led to an increase in the test score by 0.7 points (95% CI 0.3-1.1; *P*<.001).

Table 2. Anesthesia knowledge test results.

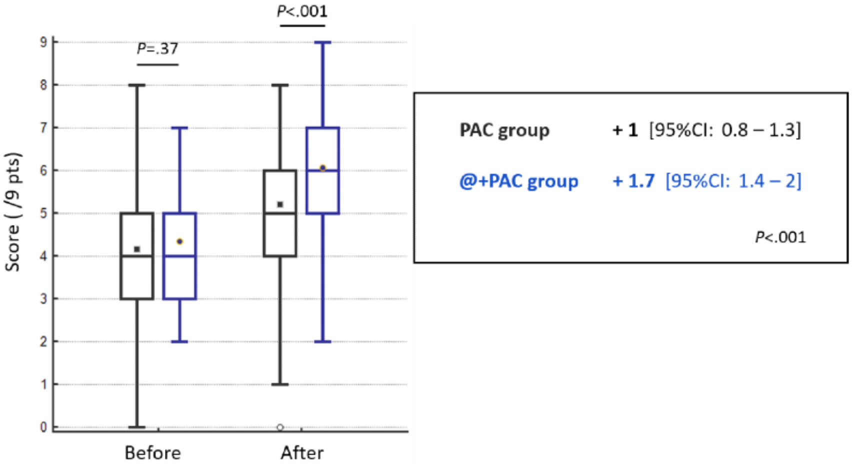
	PAC[a] group (n=205), mean (95% CI)	@+PAC[b] group (n=98), mean (95% CI)	*P* value
Knowledge test score[c] before receiving information	4.2 (3.9-4.4)	4.3 (4.0-4.7)	.37
Knowledge test score[a] after receiving information	5.2 (5.0-5.4)	6.1 (5.8-6.4)	<.001
Score improvement	+1.0 (0.8-1.3)	+1.7 (1.4-2.0)	<.001

[a] PAC: preanesthetic consultation

[b] @+PAC: digital conversational agent and preanesthetic consultation

[c] Score range is 0-9 points.

Figure 4. Comparison of anesthesia knowledge test scores between patient groups. Means are shown as markers within the box of 25th and 75th percentile values; whiskers represent ranges. PAC: preanesthetic consultation; @+PAC: digital conversational agent and preanesthetic consultation; pts: points.



Patients' Satisfaction

Among patients in the @+PAC group, 74% (73/98) agreed that the digital conversational agent was easy to access, and 82% (80/98) found the information to be clear and appropriate (Table 3).

In the PAC group, 34.1% (70/205) of patients agreed and 54.1% (111/205) disagreed with the following statement: "I wish I had received information before the anesthetic consultation."

Table 3. Digital conversational agent users' satisfaction analysis (n=98).

	Digital conversational agent accessibility, n (%)	Digital conversational agent content quality, n (%)
Number of respondents	98 (100)	98 (100)
Number of respondents who rated 1-2[a]	10 (10)	4 (4)
Number of respondents who rated 3[a]	15 (15)	14 (14)
Number of respondents who rated 4-5[a]	73 (74)	80 (82)

[a] Likert scale: 1=strongly disagree, 2=rather disagree, 3=no opinion, 4=rather agree, 5=strongly agree.

Among patients in the @+PAC group, the videos in the support and technique sections of the digital conversational agent were the most watched among viewers (88/98, 90%). The video in the section about the recovery room was the least popular, but was still watched by 68% (67/98) of the patients who logged in.

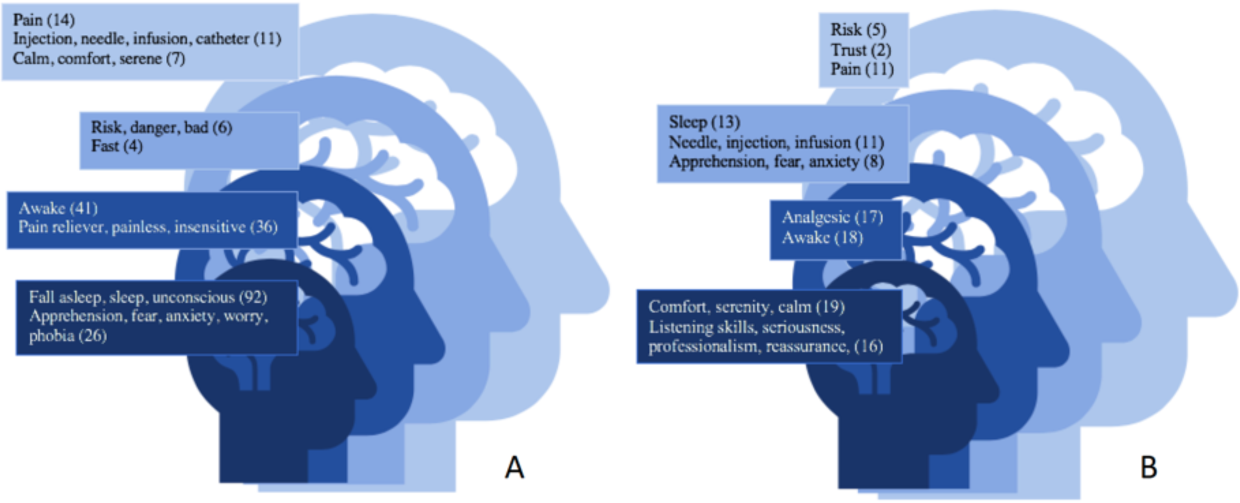
The apparent success of the videos contrasts with the low consultation rate regarding information delivered in the FAQs. Indeed, among the patients in the @+PAC group, only 4% (4/98) consulted all 11 FAQs, and 66% (65/98) did not consult any.

Patients' Representations

Responses from both groups were pooled.

Before receiving the preanesthetic information, the 233 patient responses indicated a core representation of anesthesia made up of concerns related to sedation (92 occurrences in the first row) and to apprehension (26 occurrences in the first row) (Figure 5A). This apprehension was confirmed by the elements from the first periphery, with the notion of awakening in particular (41 occurrences in the second row) associated with the problem of pain expectations (36 occurrences in the second row).

Figure 5. Graphic representation of patients' feelings before (A) and after (B) receiving information using the Abric method of hierarchized evocation. From the center to the periphery are spread out: the central core area, the first periphery, the contrasting elements, and the second periphery. The number of occurrences of the terms is indicated in parentheses.



The contrasted elements were risk and fasting, brought up in row 1, although rarely. Infrequently brought up in lower rows, pain (14 occurrences), injection (11 occurrences), and calm (7 occurrences) defined the second periphery.

After information was received, the 149 responses showed a change in the core representation, with a new interest in comfort (19 occurrences in the first row) and skills of the health care professionals (16 occurrences in the first row) (Figure 5B). Elements related to the absence of pain and the awakening no longer presented the initial strength, with 17 and 18 occurrences, respectively, in the third rows. However, elements related to injections (11 occurrences) came to the first row even if the frequency was still low; elements regarding apprehension remained in the first row but became rarer (8 occurrences). Risk, which was previously in the first row, moved to the third row and at a low frequency (5 occurrences). The notion of trust also appeared (2 occurrences in the second row).

Discussion

Using a digital conversational agent before the PAC lead to a significant improvement in the patients' knowledge of anesthesia. The high rate of internet access in addition to the high acceptability of the digital conversational agent encourage us to widely develop this tool. Our results highlight the lack of patients' knowledge about anesthesia. Indeed, no patient in either group received a passing score on the test before information delivery. In addition, the low increase in test scores (by 1 point) after PAC alone speaks in favor of identifying the best platforms and methods for delivering quality information.

Patient education through the use of information and communication technologies is a rapidly developing field that promises to improve patient outcomes while simultaneously using fewer human resources. In terms of content format, our study found that videos were consulted much more than text (ie, FAQs). Our results are in accordance with the meta-analysis conducted by Lee et al [15] in which patients being offered explanatory videos had better odds of correctly answering questions regarding anesthesia (relative risk 6.6, 95% CI 2.1-21.5). Hering et al [16] also showed an improvement in patients' satisfaction and knowledge by visiting a website before the PAC.

In our study, the most-viewed videos were related to anesthesia techniques and hospitalization modalities. Of course, these results could guide us toward themes that patients may wish to address before surgery.

Managing the content and accuracy of information offered enables better guidance with patients' online research, which may be beneficial because unsupervised online research can occasionally create anxiety [17].

Another advantage of digital tools is monitoring and re-evaluation. The data from the connections to the digital companion enable us to reliably monitor its use, allowing us to regularly readjust and update the content, unlike paper-based materials, where use cannot be evaluated.

Our results confirm that increasing patients' information is necessary to improve their satisfaction [18-21] and knowledge [16,19], as well as to reduce their anxiety [22-25].

We made the choice to place the interaction with the digital companion before the PAC to stimulate patients' curiosity and prompt them to ask themselves questions about the modes of anesthesia and their care pathways. However, surprisingly, we noticed during phase 1 that almost one-half of the patients did not wish to receive any information before the PAC. Thus, the question of information timing remains open. Interestingly, the use of a digital companion is a real advantage because it is easily accessible at any time, before and after consultation. In addition, this kind of interactive online tool could make communication with patients more efficient, especially when hospitals are not accessible, such as during a pandemic.

Giving information before a consultation could also shorten the length of the PAC without impacting patients' satisfaction [26,27]. Taylor et al [27] evaluated patients' completion of a numerical questionnaire before consulting with the anesthetist nurse. The mean consultation time of the group being offered the questionnaire was 12 minutes compared with 27 minutes for the group who was not offered the questionnaire ($P<.001$).

Other studies have shown the possibility of reliably automating the score calculation of patients' anesthetic risks [28]. Using a digital questionnaire before PAC would reduce the consultation duration while insuring a high level of quality and patient satisfaction [29].

For the first time, the method of hierarchized evocation was used to observe patients' feelings toward anesthesia. Overall, our results display patients' anxiety before receiving information and the benefits of getting information on how anesthesia is perceived. The terms skill, listening, seriousness, professionalism, comfort, and reassurance frequently appear after information is received. The semantic field category of these words is the correlate of a low preoperative anxiety level.

Our study presents several limits. First, the results were obtained by enrolling patients scheduled for elective surgery, and the results may not be able to be extrapolated to emergency cases where the consultation with a digital tool may not be feasible. Second, only 16.3% (98/600) of questionnaires distributed during phase 2 were completed. Characteristics of the nonresponders (eg, advanced age) could be very informative in identifying possible barriers. Finally, we have not evaluated physician satisfaction, which could help to identify the best perioperative health education support for the future.

To conclude, we have shown improvement in patients' knowledge about their care pathway when a digital conversational agent was used before the PAC. Despite the encouraging results, the overall uptake of the tool was relatively low and, even when used, did not achieve maximum impact. Future studies should focus on adapting both the content and the delivery of a digital conversational agent for the PAC in order to maximize its benefit to patients.

Acknowledgments

We would like to thank Olivier Thuillart and Jean-Louis Fraysse from BOTdesign (Toulouse, France), the company that created the digital companion. Support was provided solely from department sources. This work should be attributed to the Département d'Anesthésie-Réanimation, Centre Hospitalier Universitaire de Toulouse Purpan, Toulouse, France.

Authors' Contributions

FF, NB, BB, and AP performed all of the statistical analyses and wrote the manuscript. AC, FM, MK, OF, and VM participated in the design of the study and helped to draft the manuscript. FM shared his expertise in patient education. FF, AF, and LB performed all of the preanesthetic consultations. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of Frequently Asked Questions extracted from the chatbot.
[DOCX File , 992 KB - jmir_v22i12e20455_app1.docx]

Multimedia Appendix 2
Satisfaction questions assessed by the Likert scale.
[DOCX File , 15 KB - jmir_v22i12e20455_app2.docx]

References

1. Décret no 94-1054 du 5 décembre 1994 relatif aux conditions techniques de fonctionnement des établissements de santé en ce qui concerne la pratique de l'anesthésie et modifiant le code de la santé publique. Journal Officiel de La République Française. 1994 Dec 05. URL: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000549818/ [accessed 2020-11-02]
2. Ordre National des Médecins. Recommandations concernant les relations entre anesthésistes-réanimateurs et chirurgiens, autres spécialistes ou professionnels de santé. Société Française d'Anesthésie et de Réanimation 2002:1-20 [FREE Full text]
3. Ecoffey C, Pontone S. Démographie médicale en anesthésie-réanimation: causes et conséquences ? Réanimation. Conférence d'actualisation; Paris: Société Française d'Anesthésie Réanimation; 2017 Sep. URL: http://sofia.medicalistes.fr/spip/IMG/pdf/demographie_medicale_en_anesthesie_reanimation_causes_et_consequences.pdf [accessed 2020-09-30]
4. Swinhoe CF, Groves ER. Patients' knowledge of anaesthetic practice and the role of anaesthetists. Anaesthesia 1994 Feb;49(2):165-166. [doi: 10.1111/j.1365-2044.1994.tb03380.x] [Medline: 8093141]
5. Piau A, Crissey R, Brechemier D, Balardy L, Nourhashemi F. A smartphone Chatbot application to optimize monitoring of older patients with cancer. Int J Med Inform 2019 Aug;128:18-23. [doi: 10.1016/j.ijmedinf.2019.05.013] [Medline: 31160007]
6. Bibault J, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A Chatbot Versus Physicians to Provide Information for Patients With Breast Cancer: Blind, Randomized Controlled Noninferiority Trial. J Med Internet Res 2019 Nov 27;21(11):e15787 [FREE Full text] [doi: 10.2196/15787] [Medline: 31774408]
7. Macario A, Weinger M, Carney S, Kim A. Which clinical anesthesia outcomes are important to avoid? The perspective of patients. Anesth Analg 1999 Sep;89(3):652-658. [Medline: 10475299]
8. Matthey P, Finucane BT, Finegan BA. The attitude of the general public towards preoperative assessment and risks associated with general anesthesia. Can J Anaesth 2001 Apr;48(4):333-339. [doi: 10.1007/BF03014959] [Medline: 11339773]
9. Matthey PW, Finegan BA, Finucane BT. The public's fears about and perceptions of regional anesthesia. Reg Anesth Pain Med 2004;29(2):96-101. [doi: 10.1016/j.rapm.2003.10.017] [Medline: 15029543]
10. Demonstration version of the digital conversational agent MyAnesth (using password 0000). URL: http://www.my-infinity.net/orthopedie [accessed 2020-10-05]
11. Hsu C, Sandford B. The Delphi Technique: Making Sense of Consensus. Practical Assessment, Research, and Evaluation Internet. URL: https://scholarworks.umass.edu/pare/vol12/iss1/10 [accessed 2020-09-30]
12. Abric J. L'approche structurale des représentations sociales: développements récents. Psychologie et société 2001;2(4):81-104.
13. Verges P. The evocation of money: A method for defining the central core of a representation. Bulletin de psychologie 1992;45:203-209. [doi: 10.3917/bupsy]
14. Lo Monaco G, Lheureux F. Représentations sociales: théorie du noyau central et méthodes d'étude. Revue Electronique de Psychologie Sociale Internet 2007:1-55 [FREE Full text]
15. Lee A, Chui PT, Gin T. Educating patients about anesthesia: a systematic review of randomized controlled trials of media-based interventions. Anesth Analg 2003 May;96(5):1424-31, table of contents. [Medline: 12707146]
16. Hering K, Harvan J, Dangelo M, Jasinski D. The use of a computer website prior to scheduled surgery (a pilot study): impact on patient information, acquisition, anxiety level, and overall satisfaction with anesthesia care. AANA J 2005 Feb;73(1):29-33. [Medline: 15727280]
17. Tanis M, Hartmann T, Te PF. Online health anxiety and consultation satisfaction: A quantitative exploratory study on their relations. Patient Educ Couns 2016 Dec;99(7):1227-1232. [doi: 10.1016/j.pec.2016.01.021] [Medline: 26873545]
18. Kaur H, Singh G, Singh A, Sharda G, Aggarwal S. Evolving with modern technology: Impact of incorporating audiovisual aids in preanesthetic checkup clinics on patient education and anxiety. Anesth Essays Res 2016;10(3):502-507 [FREE Full text] [doi: 10.4103/0259-1162.177187] [Medline: 27746541]
19. Schenker Y, Fernandez A, Sudore R, Schillinger D. Interventions to improve patient comprehension in informed consent for medical and surgical procedures: a systematic review. Med Decis Making 2011;31(1):151-173. [doi: 10.1177/0272989X10364247] [Medline: 20357225]
20. Snyder-Ramos SA, Seintsch H, Böttiger BW, Motsch J, Martin E, Bauer M. Patient satisfaction and information gain after the preanesthetic visit: a comparison of face-to-face interview, brochure, and video. Anesth Analg 2005 Jun;100(6):1753-1758. [doi: 10.1213/01.ANE.0000153010.49776.E5] [Medline: 15920209]

21. West AM, Bittner EA, Ortiz VE. The effects of preoperative, video-assisted anesthesia education in Spanish on Spanish-speaking patients' anxiety, knowledge, and satisfaction: a pilot study. J Clin Anesth 2014 Jun;26(4):325-329. [doi: 10.1016/j.jclinane.2013.12.008] [Medline: 24882604]
22. Cheung A, Finegan BA, Torok-Both C, Donnelly-Warner N, Lujic J. A patient information booklet about anesthesiology improves preoperative patient education. Can J Anesth/J Can Anesth 2007 May;54(5):355-360. [doi: 10.1007/bf03022657]
23. Fitzgerald BM, Elder J. Will a 1-page informational handout decrease patients' most common fears of anesthesia and surgery? J Surg Educ 2008;65(5):359-363. [doi: 10.1016/j.jsurg.2008.07.013] [Medline: 18809166]
24. Jlala HA, French JL, Foxall GL, Hardman JG, Bedforth NM. Effect of preoperative multimedia information on perioperative anxiety in patients undergoing procedures under regional anaesthesia. Br J Anaesth 2010 Mar;104(3):369-374. [doi: 10.1093/bja/aeq002] [Medline: 20124283]
25. Mohan B, Kumar R, Attri JP, Chatrath V, Bala N. Anesthesiologist's Role in Relieving Patient's Anxiety. Anesth Essays Res 2017;11(2):449-452 [FREE Full text] [doi: 10.4103/0259-1162.194576] [Medline: 28663639]
26. Howell M, Hood AJ, Jayne DG. Use of a patient completed iPad questionnaire to improve pre-operative assessment. J Clin Monit Comput 2017 Feb;31(1):221-225. [doi: 10.1007/s10877-015-9818-0] [Medline: 26715416]
27. Taylor SK, Andrzejowski JC, Wiles MD, Bland S, Jones GL, Radley SC. A prospective observational study of the impact of an electronic questionnaire (ePAQ-PO) on the duration of nurse-led pre-operative assessment and patient satisfaction. PLoS One 2018;13(10):e0205439 [FREE Full text] [doi: 10.1371/journal.pone.0205439] [Medline: 30339687]
28. Zuidema X, Tromp Meesters RC, Siccama I, Houweling PL. Computerized model for preoperative risk assessment. Br J Anaesth 2011 Aug;107(2):180-185 [FREE Full text] [doi: 10.1093/bja/aer151] [Medline: 21659404]
29. Goodhart IM, Andrzejowski JC, Jones GL, Berthoud M, Dennis A, Mills GH, et al. Patient-completed, preoperative web-based anaesthetic assessment questionnaire (electronic Personal Assessment Questionnaire PreOperative): Development and validation. Eur J Anaesthesiol 2017 Apr;34(4):221-228. [doi: 10.1097/EJA.0000000000000545] [Medline: 27798453]

Abbreviations

PAC: preanesthetic consultation

Edited by G Eysenbach; submitted 19.05.20; peer-reviewed by T Aslanidis, K Wild; comments to author 16.06.20; revised version received 14.08.20; accepted 16.08.20; published 04.12.20.

Please cite as:
Ferré F, Boeschlin N, Bastiani B, Castel A, Ferrier A, Bosch L, Muscari F, Kurrek M, Fourcade O, Piau A, Minville V
Improving Provision of Preanesthetic Information Through Use of the Digital Conversational Agent “MyAnesth”: Prospective Observational Trial
J Med Internet Res 2020;22(12):e20455
URL: https://www.jmir.org/2020/12/e20455
doi:10.2196/20455
PMID:33275108

Original Paper

Readiness for Voice Technology in Patients With Cardiovascular Diseases: Cross-Sectional Study

Małgorzata Kowalska[1], PhD; Aleksandra Gładyś[2], MS; Barbara Kalańska-Łukasik[3], MD; Monika Gruz-Kwapisz[3], MD; Wojciech Wojakowski[3], MD, PhD; Tomasz Jadczyk[3,4,5], MD, PhD

[1]Department of Epidemiology, Faculty of Medical Sciences in Katowice, Medical University of Silesia, Katowice, Poland
[2]Student Scientific Society, Department of Epidemiology, Faculty of Medical Sciences in Katowice, Medical University of Silesia, Katowice, Poland
[3]Department of Cardiology and Structural Heart Diseases, Medical University of Silesia, Katowice, Poland
[4]Interventional Cardiac Electrophysiology Group, International Clinical Research Center, St Anne's University Hospital Brno, Brno, Czech Republic
[5]Research and Development Division, CardioCube Corp, Seattle, WA, United States

Corresponding Author:
Tomasz Jadczyk, MD, PhD
Department of Cardiology and Structural Heart Diseases
Medical University of Silesia
Ziolowa 45-47
Katowice, 40-635
Poland
Phone: 48 322523930
Fax: 48 323598887
Email: tomasz.jadczyk@gmail.com

Abstract

Background: The clinical application of voice technology provides novel opportunities in the field of telehealth. However, patients' readiness for this solution has not been investigated among patients with cardiovascular diseases (CVD).

Objective: This paper aims to evaluate patients' anticipated experiences regarding telemedicine, including voice conversational agents combined with provider-driven support delivered by phone.

Methods: A cross-sectional study enrolled patients with chronic CVD who were surveyed using a validated investigator-designed questionnaire combining 19 questions (eg, demographic data, medical history, preferences for using telehealth services). Prior to the survey, respondents were educated on the telemedicine services presented in the questionnaire while being assisted by a medical doctor. Responses were then collected and analyzed, and multivariate logistic regression was used to identify predictors of willingness to use voice technology.

Results: In total, 249 patients (mean age 65.3, SD 13.8 years; 158 [63.5%] men) completed the questionnaire, which showed good repeatability in the validation procedure. Of the 249 total participants, 209 (83.9%) reported high readiness to receive services allowing for remote contact with a cardiologist (176/249, 70.7%) and telemonitoring of vital signs (168/249, 67.5%). The voice conversational agents combined with provider-driven support delivered by phone were shown to be highly anticipated by patients with CVD. The readiness to use telehealth was statistically higher in people with previous difficulties accessing health care (OR 2.920, 95% CI 1.377-6.192) and was most frequent in city residents and individuals reporting a higher education level. The age and sex of the respondents did not impact the intention to use voice technology (P=.20 and P=.50, respectively).

Conclusions: Patients with cardiovascular diseases, including both younger and older individuals, declared high readiness for voice technology.

(J Med Internet Res 2020;22(12):e20456) doi:10.2196/20456

KEYWORDS
voice technology; smart speaker; acceptance; telehealth; cardiovascular diseases; chatbot

Introduction

The dynamic growth of telehealth and telemedicine solutions enables patients to access clinical care remotely, which equalizes health care coverage, supports care coordination, and increases the safety of patients and providers through physical distancing. Well-established phone and video consultations [1], along with emerging technologies [2,3], offer a powerful framework for connected health. Among patients with cardiovascular disease (CVD), telemedicine is supported by the European Society of Cardiology and the American Heart Association, which recommend multidisciplinary programs and tele-education for cardiovascular care [4-6]. Moreover, recent meta-analyses concluded that virtual care is more effective for adults with heart failure (HF) compared with standard care, particularly in reducing all-cause mortality, cardiac mortality, and hospitalization rates [7,8]. However, the effectiveness of telehealth programs depends on patient engagement and willingness to use the services [9,10]. Thus, it is crucial to understand patients' expectations and consider them while planning and developing novel telemedical strategies.

The latest advances in the field of artificial intelligence (AI) and natural language understanding have paved a way for the wide application of voice technology (VT) in health care through smart speakers and mobile apps. First real-world implementations have confirmed that AI conversational agents have the potential to support clinical care and optimize workflows [3,11-13]. Nationwide Children’s Hospital (in Columbus, Ohio) implemented voice technology for pediatric medicine [11], while Mayo Clinic developed Amazon’s Alexa software to automatically answer COVID-19–related questions [12]. Furthermore, CardioCube software (CardioCube Corp.) deployed on Amazon Alexa was shown to be functional for paperless medical history collection from patients with CVD [3] as well as for long-term remote follow-up of individuals with HF at home [13]. To evaluate prevailing perceptions and expectations, our study analyzed the opinions of and readiness for VT-based telehealth solutions among patients with CVD.

Methods

Study Design

The cross-sectional study was based on the investigator-designed and validated 19-item questionnaire used to recognize factors influencing readiness for telemedicine among individuals with CVD. The methodology included the (1) preparation of the questionnaire (Multimedia Appendix 1), (2) assessment of the questionnaire’s repeatability (questionnaire validation process), and (3) evaluation of anticipated patient experiences regarding telemedical solutions using the previously validated questionnaire, which was provided to the total studied population.

This study was performed at the Department of Cardiology and Structural Heart Diseases, Medical University of Silesia in Katowice, Poland, which hospitalizes approximately 2700 to 2800 individuals per year. The ethics approval for the study was received from the Bioethical Committee of the Medical University of Silesia in Katowice (approval number KNW/0022/KB1/160/1617) on February 3, 2017.

Questionnaire Validation

The original questionnaire included 19 questions concerning demographic data (sex, age, place of residence, educational level, occupational activity); health status data, such as previously diagnosed diseases, history of cardiovascular treatments, and previous hospitalization; and information about the current form of follow-up contact with a cardiologist. Key questions were related to the patient’s opinions and personal preferences regarding the application of telemedicine for remote contact with a medical doctor. To ensure validity of the questionnaire, the same 30 patients answered the questionnaire twice within the 2- to 3-day interval. The median age of the participants was 65 years. Most of the respondents were men (21/30, 70%), lived in the city (26/30, 87%), had a higher level of education (12/30, 40%), and reported living with another family member (23/30, 77%). To assess the reliability of the questionnaire, the Cohen statistic and interclass correlation coefficient were calculated with simultaneous assessment of the percentage of the repeatability of responses for the key questions of the questionnaire. By the applicable principles, the mean, good, and very good compliance of the Cohen values are represented by the values of 0.41 to 0.60, 0.61 to 0.80, and 0.81 to 1.00, respectively [14]. The key questions had very high repeatability, ranging from 80% to 100%. Detailed results of the validation procedure are presented in Multimedia Appendix 2. As confirmed by the research tool reliability, study participants were surveyed using a validated author-created questionnaire.

Cross-Sectional Study

Between March 2019 and January 2020, each of the patients admitted to the Department of Cardiology was invited to participate in the questionnaire survey. Written consent to participate in the study was obtained from 249 patients (participation rate of 9%). The participants completed the questionnaires while assisted by the medical doctor, who explained the meaning and potential clinical application of the telemedical solutions presented in the survey, assuring that responders understood the questions (Multimedia Appendix 3).

Statistical Analysis

Statistical analysis was undertaken using the Statistica 13.0 package (Dell Software). The missing values were removed pairwise from further statistical analysis. The measures of central tendency (median, quartile) and dispersion (interquartile range) were applied in the statistical description of quantitative variables; their distribution was verified using the Shapiro-Wilk test. Qualitative variables were presented using frequency and percentage. Differences between groups were tested using the chi-square or Fisher test. In all analyses, *P* values below .05 were considered statistically significant. Finally, the relevant relationships between particular variables were verified using multivariable analysis (logistic regression models with Hooke-Jeeves and quasi-Newton estimation). The null hypothesis in the adopted model was that patients’ readiness to use telemedicine solutions in general and to use a voice conversational agent specifically are not related to independent parameters, such as female sex, age, place of residence being in the city, primary level of education, living with family, and previous difficulties with medical care. Only the statistically significant variables obtained in bivariate analyses were included in the model. Additionally, a chi-square test was used to evaluate the goodness of fit of the logistic regression models. A *P* value above .05 indicates a good fit of the model.

Results

Participant Demographics

The total studied group included 249 patients (158/249, 63.5% men) aged 65.3 (SD 13.8) years. Table 1 presents basic characteristics of the studied population.

Hypertension, atherosclerosis, and heart failure were the most common diseases diagnosed in the surveyed population. More than half of the respondents (144/249, 57.8%) were hospitalized approximately twice in the previous year. Coronary angiography (146/249, 58.6%) and percutaneous coronary interventions (105/249, 42.2%) were reported as the most frequent medical procedures (Multimedia Appendix 4). In terms of health care access, approximately half of patients (125/249, 50.2%) obtained medical services at public outpatient clinics. Almost every tenth person (23/249, 9.2%) used only private clinics, while every third patient (90/249, 36.1%) scheduled follow-up visits both at private and public medical centers.

Table 1. Baseline socioeconomic characteristics.

Qualitative variables	Participants, n (%)
Sex	
Male	158 (63.5)
Female	91 (36.5)
Place of residence	
City	211 (84.7)
Rural	38 (15.3)
Marital status	
In relationship	193 (77.5)
Living alone	53 (21.3)
No data	3 (1.2)
Educational level	
Primary	111 (44.6)
Secondary	75 (30.1)
Higher	63 (25.3)
Socioeconomic activity	
Retired	167 (67.1)
Occupationally active	61 (24.5)
Student	6 (2.4)
No data	15 (6.0)
Phone access	
Yes	246 (98.8)
No	3 (1.2)
Internet access	
Yes	158 (63.5)
No	91 (36.5)

Readiness for Telemedicine Solutions

The vast majority of respondents (228/249, 91.6%) remained under ambulatory follow-up care, with periodic visits to an outpatient cardiology clinic. Consistently, of the 249 total participants, 209 (83.9%) said they would accept telemedicine solutions, while 34 (13.6%) were definitely against telemedicine solutions and 6 did not respond. The results confirmed that nearly three-fourths of the respondents expressed acceptance of telemedicine to maintain contact with a physician as well as for remote monitoring of vital signs. Further details are presented in Table 2.

Furthermore, the results showed that mobile phones (167/249, 67.1%) and VT combined with provider-driven phone support (166/249, 66.7%) were the preferred communication channels between patients and doctors. Furthermore, patients' preferences regarding the envisioned form of care showed that landline phones, email contact, and webpages were chosen by 43.8% (109/249), 17.3% (43/249), and 9.2% (23/249) of participants, respectively. Of note, almost every fifth respondent chose traditional face-to-face contact with a physician (47/249, 18.9%). A total of 34 out of 249 (13.6%) patients reported that the form of contact with the doctor does not matter.

Table 2. Patients' readiness for telemedicine solutions.

Anticipated telemedicine services	Participants, n (%)
Remote contact with a cardiologist	176 (70.7)
Telemonitoring of vital signs (blood pressure, temperature, body weight)	168 (67.5)
Issuing e-prescriptions	161 (64.7)
Alarming health status deterioration	154 (61.8)
Scheduling and managing of medical visits	143 (57.4)
Medication reminder	106 (42.6)

Factors Influencing Readiness for Telemedicine Solutions

Declared readiness for receiving the presented telemedical solutions significantly depended on age, gender, previous difficulties in contacting a medical doctor, and living with family (Table 3).

Table 3. Factors influencing patients' acceptance of telemedicine.

Independent variable	Technology accepted by patients			
	Telemedicine tools, n (%)	*P* value	Voice technology combined with provider-driven phone support, n (%)	*P* value
Sex		.006		.50
Male	142 (90.4)		113 (81.9)	
Female	66 (77.6)		52 (78.8)	
Previous difficulties accessing cardiologist		<.001		.003
Yes	123 (93.9)		108 (87.8)	
No	85 (77.9)		58 (71.6)	
Living with family		.004		.40
Yes	170 (89.5)		135 (81.8)	
No	37 (74.0)		30 (76.9)	
Age		.05		.20
Younger (<68 years)	111 (90.2)		89 (84.0)	
Older (68+ years)	98 (81.7)		77 (77.8)	
Place of residence		.60		.07
City	177 (86.3)		145 (82.9)	
Rural	31 (83.8)		20 (68.9)	
Educational level		.10		.05
Primary	88 (81.5)		64 (73.6)	
Secondary	62 (88.6)		56 (88.9)	
Higher	57 (90.5)		46 (83.6)	
Internet access		.10		.30
Yes	138 (88.5)		112 (82.9)	
No	68 (81.9)		52 (77.6)	

Over half of patients (133/249, 53.4%) reported having encountered previous difficulties while contacting a doctor during routine clinical care management. The main reasons reported by study participants were long waiting times for a visit to an outpatient clinic (111/249, 44.6%), long waiting time in a clinic due to queue lengths (68/249, 27.3%), and substantial distance and travel time from their place of residence to a clinic (61/249, 24.5%). Results of the multivariate analysis confirmed that male sex, living with other family members, and previous difficulty of contact with physicians resulted in a willingness to use telemedicine solutions (Tables 4 and 5).

Table 4. Multivariate analysis between declared lack of telemedicine solution acceptance and particular independent values[a].

Independent value (n=236)	β coefficient	OR[b] (95% CI)
Intercept	–4.359	0.013 (0.000-0.570)
Sex (female)	–0.878	0.415 (0.183-0.943)
Age	0.009	1.009 (0.974-1.046)
Place of residence (city)	0.369	1.447 (0.487-4.304)
Level of education (primary)	–0.275	0.760 (0.443-1.304)
Living with family (yes)	0.894	2.445 (1.021-5.856)
Previous difficulties accessing cardiologist (yes)	1.364	3.913 (1.625-9.426)

[a] Model: logistic regression (logit); χ^2=27.1; *P*<.001.

[b] OR: odds ratio.

Table 5. Multivariate analysis between declared lack of acceptance of voice technology combined with provider-driven phone support and particular independent values[a].

Independent value (n=200)	β coefficient	OR[b] (95% CI)
Intercept	–3.408	0.033 (0.001-0.877)
Sex (female)	–0.086	0.918 (0.415-2.032)
Age	–0.001	0.999 (0.971-1.027)
Place of residence (city)	0.843	2.324 (0.879-6.143)
Level of education (primary)	–0.335	0.716 (0.436-1.175)
Living with family (yes)	0.157	1.170 (0.451-3.034)
Previous difficulties accessing cardiologist (yes)	1.072	2.920 (1.377-6.192)

[a] Model: logistic regression (logit); χ^2=13.9; *P*=.03.

[b] OR: odds ratio.

Factors Influencing Readiness for Voice Technology

The readiness for VT was statistically significantly higher among patients with previous negative experience accessing health care (odds ratio [OR] 2.920, 95% CI 1.377-6.192) and was most frequent in patients reporting higher education and in city residents (Tables 4 and 5). The age and sex of respondents did not impact the intention to use voice agents (*P*=.20 and *P*=.50, respectively) (Table 3).

The Null Hypothesis

The study results showed evidence against the null hypothesis, as the readiness to use telehealth services and, specifically, voice technology was associated with independent variables (Tables 3-5). Patients' willingness to apply telemedicine was related to gender (*P*=.006), with women being less likely to declare readiness for telehealth (OR 0.415, 95% CI 0.183-0.943) than men; age (*P*=.05), with patients younger than 68 years being more likely to receive telemedicine (OR 1.009, 95% CI 0.974-1.046); living with family (*P*=.004), with individuals living with other family member declaring higher readiness for telemedicine (OR 2.445, 95% CI 1.021-5.856); and previous difficulties accessing medical care (*P*<.001), with respondents who experienced obstacles to accessing a cardiologist in the past declaring a higher intention to use telemedicine (OR 3.913, 95% CI 1.625-9.426). Place of residence and primary level of education were not associated with readiness for telehealth (*P*=.60 and *P*=.10, respectively).

Positive responses toward voice conversational agents were associated with previous difficulties accessing cardiovascular care (*P*=.003), with individuals who encountered problems obtaining medical care in the past showing higher intention to use voice technology (OR 2.920, 95% CI 1.377-6.192), and education level (*P*=.05), with respondents with a primary education being less interested in the application of conversational agents (OR 0.716, 95% CI 0.436-1.175). Place of residence appeared to be related to the respondent's declared readiness for VT, but it did not reach statistical significance (*P*=.07). Sex, age, and living with family members were not associated with readiness for conversational agents (*P*=.50, *P*=.20, and *P*=.40, respectively).

Discussion

Principal Findings

This cross-sectional study evaluated factors influencing intention to use telemedical solutions. Survey results confirmed that patients with CVD declared readiness for remote health care services, including VT-based virtual care. Previous difficulties accessing cardiologists in routine clinical settings were associated with a 3 times higher likelihood of acceptance of voice assistants for medical purposes.

An increasing number of patients with CVD presents challenges to health care systems to provide ubiquitous and high-quality care [15-17]. Thus, the implementation of telehealth services may enable alternative treatment strategies [18]. Accordingly, telemedicine provides a broad spectrum of possible interventions, including self-management programs [19,20], medication adherence [21,22], monitoring of vital signs [23], medical visit reminders [24], and remote long-term follow-up [7]. However, the applicability and usefulness of new technologies must be viewed from a broader perspective that includes patients’ expectations. Of note, users’ judgments and opinions of services or products reflect dynamic changes in emotions and feelings over time. Anticipated experience refers to the expectations a person has prior to the use of a solution, while momentary experience is associated with a perception after the first interaction. As a continuum, episodic and remembered experience mirror users’ approaches toward the specific usage of a service and their general impressions after a longer period of use, respectively [25]. It is important to apply a systematic methodology while exploring patients' needs. The structured approach is a cornerstone for the creation of patient-oriented telehealth programs promoting high adoption rates [26].

This study evaluating anticipated experience indicated that a vast majority (123/249, 93.9%) of CVD patients with previous difficulties accessing traditional care declared a readiness to use telemedical solutions, including voice assistants. The survey revealed that almost half of respondents experienced delayed medical appointments (eg, long waiting times, queues) and difficulties with physical access associated with the distance between patients' homes and medical centers. In contrast, Edwards et al [27] showed that a relatively low number of patients indicated difficulty in accessing traditional health care as a reason to accept telehealth solutions.

Our study highlighted some sociodemographic variables impacting patients’ approaches toward anticipated telehealth opportunities. Men and individuals living with family reported higher interest in the use of telemedicine. Similarly, men were more likely to accept digital technologies in a study evaluating the adoption of health care apps and online tools [28]. The age of the patients has been reported to play an important role in patients' willingness to use telemedicine [29]. Despite our study showing a correlation between age and readiness for telehealth, it did not reach statistical significance, as 90.2% (111/123) of younger (<68 years) and 81.7% (98/120) of older (68+ years) respondents were interested in using telemedicine tools (*P*=.05). Similarly, a systematic review of 39 studies evaluating age-related acceptance of telemedicine showed no consistent association [30]. Moreover, a recent meta-analysis confirmed that different age groups expressed a specific level of mobile health service interest that depended on ease of use, perceived severity, and perceived vulnerability, especially for middle-aged and older individuals [31]. Interestingly, conversational agents were reported to be anticipated by younger and older adults (*P*=.20), which may help to avoid technological exclusion of older individuals. These findings are promising, as rapid progress and market penetration of voice assistants (ie, Amazon Echo, Google Home) mirror general users’ acceptance of verbal communication interfaces. In the United States, more than 87 million people were reported to use smart speakers in January 2020, which gives potential to the horizontal implementation of this novel telehealth modality [32]. Simultaneously, professional medical applications deployed using conversational agents allow patients to report health status by answering a set of clinical questions in the form of a verbal conversation between human and voice device. This trend is exemplified by the CardioCube application, which has already been tested [3] and implemented in clinical practice for remote monitoring of patients with HF and diabetes [13]. Institutionally, Nationwide Children’s Hospital in Columbus, Ohio, developed a voice service for the care coordination of children with medical complexities [11], which confirms the usability of VT in the routine management of patients.

Furthermore, in line with the findings reported by Lin et al [33], our results showed that patients living with their families showed higher intention to use the presented telehealth solutions. This observation might be associated with additional support from family members while using telehealth services.

It is important to note that patients declared different forms of preferred virtual care, and subsequent analysis showed a willingness to use mobile phones (167/249, 67.1% of respondents) as well as voice conversational agents (166/249, 66.7% of patients). Our results show that the coapplication of VT and provider-driven support through telephone contact meets patients’ expectations. These findings may help to create a basis for telehealth programs providing comprehensive real-time feedback on a patient's health status.

It must be noted that the psychological aspects related to the perception of new technologies are very complex. Our questionnaire focused on the critical elements of determining anticipated experience and readiness to use telehealth. We evaluated (1) reasons patients want to use remote care, (2) the biggest value propositions, (3) the types of patients most interested in using telehealth, and (4) the manner in which patients want to access care. However, there are more components associated with willingness and readiness that were not evaluated by our survey. Other studies have reported that the following factors should be taken into consideration when designing telemedicine programs: previous experience with technology, ability to use the technology, perceived usefulness, quality of design, and user confidence [20,28,34-36]. The last element is of special importance, as previous findings showed that patients trust telehealth services if they are provided directly by a physician [28] with the support of a case manager [33]. Furthermore, studies showed that factors associated with technology itself (performance expectancy and perceived privacy) determine older people's intentions to receive telehealth [37], while readiness might be also facilitated by coexisiting diseases and additional costs [38].

Study Limitations

We would like to note that we did not collect data about previous experience with telemedical solutions, which does not allow us to evaluate acceptance of virtual care. Accordingly, the study results reflected patients' anticipated experiences and intentions toward services described by a medical doctor prior to the survey. To reduce risk of bias, we used a semistructured interview to educate study participants about each of the telehealth services included in the questionnaire. Moreover, a dedicated doctor (BK) assisted all study participants during the survey, answering additional questions regarding the clinical application of telemedicine.

The study is limited by a relatively small number of participants (N=249). However, differences in answers provided by participants had strong statistical significance supporting our findings. Moreover, the validation procedure included 30 respondents, and the e-prescription question showed the lowest Cohen score (patients' responses should be taken with caution).

Taking into consideration that the study design was based on readiness for telemedicine, further investigations are needed to analyze acceptance of VT in clinical practice, including patients' episodic and remembered experiences.

Conclusion

Understanding challenges and barriers associated with the clinical use of telehealth is necessary for its successful and widespread implementation. Among many aspects, patients' expectations and factors influencing readiness play an important role. Our study showed that intention to use VT is associated with previous difficulties accessing health care, especially in city residents and individuals who reported a higher level of education. A telehealth service combining conversational agents with provider-based phone support was anticipated by younger and older adults, which may help provide effective remote management of older individuals. Further evaluation of patients' perceptions of and incentives to use remote health care technologies is critical to designing a patient-centered solution.

Conflicts of Interest

WW is an angel investor at CardioCube Corp.

Multimedia Appendix 1
Investigator-developed questionnaire.
[DOCX File , 24 KB - jmir_v22i12e20456_app1.docx]

Multimedia Appendix 2
Supplementary Table 1.
[DOCX File , 16 KB - jmir_v22i12e20456_app2.docx]

Multimedia Appendix 3
Semistructured interview about the clinical application of telemedicine.
[DOCX File , 16 KB - jmir_v22i12e20456_app3.docx]

Multimedia Appendix 4
Supplementary Table 2.
[DOCX File , 14 KB - jmir_v22i12e20456_app4.docx]

References

1. Di Lenarda A, Casolo G, Gulizia M, Aspromonte N, Scalvini S, Mortara A, et al. The future of telemedicine for the management of heart failure patients: a Consensus Document of the Italian Association of Hospital Cardiologists (A.N.M.C.O), the Italian Society of Cardiology (S.I.C.) and the Italian Society for Telemedicine and eHealth (Digital S.I.T.). Eur Heart J Suppl 2017 May;19(Suppl D):D113-D129 [FREE Full text] [doi: 10.1093/eurheartj/sux024] [Medline: 28751839]
2. Sim I. Mobile Devices and Health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/nejmra1806949]
3. Jadczyk T, Kiwic O, Khandwalla RM, Grabowski K, Rudawski S, Magaczewski P, et al. Feasibility of a voice-enabled automated platform for medical data collection: CardioCube. Int J Med Inform 2019 Sep;129:388-393. [doi: 10.1016/j.ijmedinf.2019.07.001] [Medline: 31445282]
4. Piotrowicz R, Krzesiński P, Balsam P, Kempa M, Główczyńska R, Grabowski M, et al. [Cardiology telemedicine solutions - opinion of the experts of the Committee of Informatics and Telemedicine of Polish Society of Cardiology, Section of

Non-invasive Electrocardiology and Telemedicine of Polish Society of Cardiology and Clinical Sciences C]. Kardiol Pol 2018;76(3):698-707 [FREE Full text] [doi: 10.5603/KP.a2018.0058] [Medline: 29441511]

5. Schwamm LH, Chumbler N, Brown E, Fonarow GC, Berube D, Nystrom K, et al. Recommendations for the Implementation of Telehealth in Cardiovascular and Stroke Care: A Policy Statement From the American Heart Association. Circulation 2017 Feb 14;135(7):e24-e44. [doi: 10.1161/cir.0000000000000475]
6. Ponikowski P, Voors A, Anker S, Bueno H, Cleland J, Coats A, ESC Scientific Document Group. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016 Jul 14;37(27):2129-2200. [doi: 10.1093/eurheartj/ehw128] [Medline: 27206819]
7. Zhu Y, Gu X, Xu C. Effectiveness of telemedicine systems for adults with heart failure: a meta-analysis of randomized controlled trials. Heart Fail Rev 2020 Mar 24;25(2):231-243 [FREE Full text] [doi: 10.1007/s10741-019-09801-5] [Medline: 31197564]
8. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015;10:CD007228. [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
9. Lorenz A, Oppermann R. Mobile health monitoring for the elderly: Designing for diversity. Pervasive Mobile Comput 2009 Oct;5(5):478-495 [FREE Full text] [doi: 10.1016/j.pmcj.2008.09.010]
10. Prescher S, Deckwart O, Winkler S, Koehler K, Honold M, Koehler F. Telemedical care: feasibility and perception of the patients and physicians: a survey-based acceptance analysis of the Telemedical Interventional Monitoring in Heart Failure (TIM-HF) trial. Eur J Prev Cardiol 2013 Jun;20(2 Suppl):18-24. [doi: 10.1177/2047487313487483e] [Medline: 23702987]
11. Sezgin E, Noritz G, Elek A, Conkol K, Rust S, Bailey M, et al. Capturing At-Home Health and Care Information for Children With Medical Complexity Using Voice Interactive Technologies: Multi-Stakeholder Viewpoint. J Med Internet Res 2020 Feb 13;22(2):e14202 [FREE Full text] [doi: 10.2196/14202] [Medline: 32053114]
12. Furst J. Mayo Clinic introduces skill for Amazon's Alexa about COVID-19. Mayo Clinic. 2020 Apr 27. URL: https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-introduces-skill-for-amazons-alexa-about-covid-19/ [accessed 2020-07-22]
13. CardioCube. FCNcare. URL: https://fcncare.com/ [accessed 2020-07-22]
14. Kowalska M, Baranski K. Biostatistics. Practical methods for assessing the reliability of questionnaire surveys. Katowice, Poland: The Medical University of Silesia in Katowice; 2017.
15. Benjamin E, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association. Circulation 2019 Mar 05;139(10):e56-e528. [doi: 10.1161/cir.0000000000000659]
16. Pinaire J, Azé J, Bringay S, Cayla G, Landais P. Hospital burden of coronary artery disease: Trends of myocardial infarction and/or percutaneous coronary interventions in France 2009-2014. PLoS One 2019;14(5):e0215649 [FREE Full text] [doi: 10.1371/journal.pone.0215649] [Medline: 31048833]
17. Jackson SL, Tong X, King RJ, Loustalot F, Hong Y, Ritchey MD. National Burden of Heart Failure Events in the United States, 2006 to 2014. Circ Heart Failure 2018 Dec;11(12):e004873. [doi: 10.1161/circheartfailure.117.004873]
18. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
19. Lorig KR, Ritter PL, Laurent DD, Plant K. Internet-based chronic disease self-management: a randomized trial. Med Care 2006 Nov;44(11):964-971. [doi: 10.1097/01.mlr.0000233678.80203.c1] [Medline: 17063127]
20. Kim JY, Wineinger NE, Steinhubl SR. The Influence of Wireless Self-Monitoring Program on the Relationship Between Patient Activation and Health Behaviors, Medication Adherence, and Blood Pressure Levels in Hypertensive Patients: A Substudy of a Randomized Controlled Trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
21. Quilici J, Fugon L, Beguin S, Morange PE, Bonnet J, Alessi M, et al. Effect of motivational mobile phone short message service on aspirin adherence after coronary stenting for acute coronary syndrome. Int J Cardiol 2013 Sep 20;168(1):568-569. [doi: 10.1016/j.ijcard.2013.01.252] [Medline: 23462636]
22. Khonsari S, Subramanian P, Chinna K, Latif LA, Ling LW, Gholami O. Effect of a reminder system using an automated short message service on medication adherence following acute coronary syndrome. Eur J Cardiovasc Nurs 2015 Apr;14(2):170-179. [doi: 10.1177/1474515114521910] [Medline: 24491349]
23. Blasco A, Carmona M, Fernández-Lozano I, Salvador CH, Pascual M, Sagredo PG, et al. Evaluation of a Telemedicine Service for the Secondary Prevention of Coronary Artery Disease. J Cardiopulmonary Rehabil Prev 2012;32(1):25-31. [doi: 10.1097/hcr.0b013e3182343aa7]

24. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013 Dec 05;12:CD007458 [FREE Full text] [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]
25. Marti P, Iacono I. Anticipated, momentary, episodic, remembered: the many facets of User eXperience. 2016 Presented at: Federated Conference on Computer Science and Information Systems (FedCSIS); Sep 11-14, 2016; Gdansk, Poland p. 11-14. [doi: 10.15439/2016f302]
26. Abdekhoda M, Ahmadi M, Dehnad A, Hosseini AF. Information Technology Acceptance in Health Information Management. Methods Inf Med 2018 Jan 20;53(01):14-20. [doi: 10.3414/me13-01-0079]
27. Edwards L, Thomas C, Gregory A, Yardley L, O'Cathain A, Montgomery AA, et al. Are people with chronic diseases interested in using telehealth? A cross-sectional postal survey. J Med Internet Res 2014 May;16(5):e123 [FREE Full text] [doi: 10.2196/jmir.3257] [Medline: 24811914]
28. Safi S, Danzer G, Schmailzl KJ. Empirical Research on Acceptance of Digital Technologies in Medicine Among Patients and Healthy Users: Questionnaire Study. JMIR Hum Factors 2019 Nov 29;6(4):e13472 [FREE Full text] [doi: 10.2196/13472] [Medline: 31782741]
29. Harst L, Lantzsch H, Scheibe M. Theories Predicting End-User Acceptance of Telemedicine Use: Systematic Review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
30. Or CKL, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
31. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. Int J Inform Manage 2018 Dec;43:342-350 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2017.08.006]
32. Kinsella B. Nearly 90 Million U.S. Adults Have Smart Speakers, Adoption Now Exceeds One-Third of Consumers. Voicebot. URL: https://voicebot.ai/2020/04/28/nearly-90-million-u-s-adults-have-smart-speakers-adoption-now-exceeds-one-third-of-consumers/ [accessed 2020-05-15]
33. Lin Y, Huang G, Ho Y, Lou M. Patient willingness to undergo a two-week free trial of a telemedicine service for coronary artery disease after coronary intervention: A mixed-methods study. J Nurs Manag 2020 Mar;28(2):407-416. [doi: 10.1111/jonm.12942] [Medline: 31891424]
34. Jiang J, Zhu Q, Zheng Y, Zhu Y, Li Y, Huo Y. Perceptions and Acceptance of mHealth in Patients With Cardiovascular Diseases: A Cross-Sectional Study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e10117 [FREE Full text] [doi: 10.2196/10117] [Medline: 30714942]
35. Lee J, Nguyen AL, Berg J, Amin A, Bachman M, Guo Y, et al. Attitudes and preferences on the use of mobile health technology and health games for self-management: interviews with older adults on anticoagulation therapy. JMIR Mhealth Uhealth 2014 Jul 23;2(3):e32 [FREE Full text] [doi: 10.2196/mhealth.3196] [Medline: 25098413]
36. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht U. Mobile technologies: expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR Mhealth Uhealth 2014 Oct 21;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
37. van Houwelingen CT, Ettema RG, Antonietti MG, Kort HS. Understanding Older People's Readiness for Receiving Telehealth: Mixed-Method Study. J Med Internet Res 2018 Apr 06;20(4):e123 [FREE Full text] [doi: 10.2196/jmir.8407] [Medline: 29625950]
38. Muigg D, Duftschmid G, Kastner P, Modre-Osprian R, Haluza D. Telemonitoring readiness among Austrian diabetic patients: A cross-sectional validation study. Health Informatics J 2020 Feb 12:E [FREE Full text] [doi: 10.1177/1460458219894094] [Medline: 32046567]

Abbreviations

AI: artificial intelligence
CVD: cardiovascular disease
HF: heart failure
OR: odds ratio
VT: voice technology

Edited by G Eysenbach; submitted 19.05.20; peer-reviewed by E Sezgin, E Bellei; comments to author 08.06.20; revised version received 21.07.20; accepted 17.09.20; published 17.12.20.

Please cite as:
Kowalska M, Gładyś A, Kalańska-Łukasik B, Gruz-Kwapisz M, Wojakowski W, Jadczyk T
Readiness for Voice Technology in Patients With Cardiovascular Diseases: Cross-Sectional Study
J Med Internet Res 2020;22(12):e20456
URL: http://www.jmir.org/2020/12/e20456/
doi:10.2196/20456
PMID:33331824

Original Paper

A Blockchain-Based Consent Platform for Active Assisted Living: Modeling Study and Conceptual Framework

Pedro Elkind Velmovitsky[1], BSc, MSc; Pedro Augusto Da Silva E Souza Miranda[1], BSc, MSc; Hélène Vaillancourt[2], PEng, PhD; Tania Donovska[2], BSc, MEng, PMP; Jennifer Teague[2], PhD; Plinio Pelegrini Morita[1,3,4,5,6], PEng, MSc, PhD

[1]School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada
[2]CSA Group, Toronto, ON, Canada
[3]Research Institute for Aging, University of Waterloo, Waterloo, ON, Canada
[4]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada
[5]eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada
[6]Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Plinio Pelegrini Morita, PEng, MSc, PhD
School of Public Health and Health Systems
University of Waterloo
200 University Ave W
Waterloo, ON, N2L 3G1
Canada
Phone: 1 5198884567 ext 31372
Email: plinio.morita@uwaterloo.ca

Abstract

Background: Recent advancements in active assisted living (AAL) technologies allow older adults to age well in place. However, sensing technologies increase the complexity of data collection points, making it difficult for users to consent to data collection. One possible solution for improving transparency in the consent management process is the use of blockchain, an immutable and timestamped ledger.

Objective: This study aims to provide a conceptual framework based on technology aimed at mitigating trust issues in the consent management process.

Methods: The consent management process was modeled using established methodologies to obtain a mapping of trust issues. This mapping was then used to develop a conceptual framework based on previous monitoring and surveillance architectures for connected devices.

Results: In this paper, we present a model that maps trust issues in the informed consent process; a conceptual framework capable of providing all the necessary underlining technologies, components, and functionalities required to develop applications capable of managing the process of informed consent for AAL, powered by blockchain technology to ensure transparency; and a diagram showing an instantiation of the framework with entities comprising the participants in the blockchain network, suggesting possible technologies that can be used.

Conclusions: Our conceptual framework provides all the components and technologies that are required to enhance the informed consent process. Blockchain technology can help overcome several privacy challenges and mitigate trust issues that are currently present in the consent management process of data collection involving AAL technologies.

(J Med Internet Res 2020;22(12):e20832) doi:10.2196/20832

KEYWORDS

health care; blockchain; Internet of Things; aging; informed consent; mobile phone; computing methodologies; computer security

Introduction

Background

Society is currently moving into an age of ubiquitous and smart technologies, including mobile and wearable products, portable sensors, and diverse internet of things (IoT) solutions [1]. Examples of such smart technologies are smartphones, smartwatches, wireless blood pressure cuffs, wireless scales, smart thermostats, and smart homes, among others [2-4]. These devices have had a substantial adoption rate; for example, in 2016, 76% of Canadians owned a smartphone [5]. These devices have become a standard and are pervasive even in developing countries; for example, in Brazil, 57% of the population used a smartphone [6], and in Argentina, a similar proportion of the population, that is, 52%, used a smartphone [7]. According to a recent survey of 11 developing countries across 4 global regions, a median of 53% of the population of these nations have access to a smartphone with internet [8]. The wearable market has also experienced rapid growth: the number of global smartwatch users increased from 5 million in 2014 to 75 million in 2017 [9]. The fitness wearable market alone, comprising devices such as fitness bands, smart clothes, and eyewear (eg, smart glasses), has approximately 4 million users in Canada with a revenue of Can $ 290 (US $220) million [10].

These technologies have embedded sensors that can continuously and effortlessly monitor the health of users [11] by collecting data on vital signs [12], environmental variables [13,14], and behavioral metrics such as movement in the house [1]. The collected data can be analyzed for new insights into the health of individuals and populations [15].

Active Assisted Living

One of the fields that IoT technologies and connected devices can greatly improve and support is active assisted living (AAL) [16], defined as "all technology, devices, and wearables connected to the Internet, that enable the collection and exchange of data, and are used for health monitoring or to enhance the daily life of individuals" [17]. The major goal of AAL technology is to help people with activities of daily living, leading them to a better, safer, and more productive life while minimizing the risk of injury and avoidable death [1,17]. Therefore, AAL technologies are usually designed to support vulnerable populations and older adults [1,17].

AAL has become increasingly important because of a shift toward older populations in the global age distribution. It is expected that by 2050, all developing countries, in addition to Latin America, the Caribbean, and most of Asia, will have a median age of at least 40 years [18]. In Canada, 10 years from now, for example, 1 out of 4 people will be aged over 65 years [19].

Aging well and in place requires the older adult population to be as independent, secure, and healthy as possible [1,17]. However, older adults experience declining health as they age and are more likely to develop some sort of impairment, making remaining at home a difficult task [20]. Although assistive technologies have long been used to help older adults with daily activities [20], the benefits of innovative connected devices for monitoring wellness and supporting aging-in-place are being realized only recently. Their sensing capabilities allow for health support, real-time data collection, and detection or warning for emergencies. For example, 1 out of 4 seniors fall each year, resulting in the deaths of 27,000 seniors every year in the United States alone [21]. Another example of how these technologies can support older adults is a smart thermometer that cools the house if the temperature increases to dangerous levels, preventing seniors from experiencing a heatstroke [1].

Controlling access and consent to all these data is still a great challenge in the current technological landscape [18,22-26].

Informed Consent

Informed consent is defined as an "individual's autonomous authorization of medical intervention or participation in research" [18]. Valid informed consent in research and treatment is composed of the following minimum requirements [18,27]:

1. Competence: defined as the ability to perform a task
2. Understanding: defined as the full disclosure of information pertaining to the situation
3. Voluntariness: defined as participation in the event without any coercion and awareness of the possible outcomes.

Additional complexities in the consent management process may arise depending on the case; for example, older adults use AAL systems tend to interact with caregivers, cohabitants, and legal guardians. AAL technologies, such as smart thermostats, will collect data from every person in the environment where they are located. Therefore, the caregivers, cohabitants, and legal guardians must also give consent for data collection and use [18,23].

In addition, AAL technologies have significant benefits for people with cognitive impairments, including older adults and people with dementia [18]. A delicate ethical consideration arises on whether these populations can provide valid informed consent based on the principles outlined above [18]. One solution would be *rolling informed consent*, where the data collector repeatedly provides information on consent while assessing the mental capability of an individual [18,28]. Although rolling informed consent may work well in research environments, in practice, it may not be possible to achieve it in a real-world deployment of AAL systems because of the high complexity, rate, and volume of data collection from devices.

In cases where the individuals whose data are being collected cannot give valid informed consent, substitute decision makers (SDMs) may be necessary to make health decisions on their behalf [29]. SDMs may be family members, caregivers, legal guardians, or any person authorized to make health decisions for or with the individual.

Privacy of AAL Data

Despite the benefits to technology users, health care providers, and the health care system, one major challenge that needs to be addressed is the privacy of patient-generated health data. Although IoT devices allow continuous and zero-effort monitoring of health data, they also increase the complexity of data collection points and make it harder to determine exactly what, why, and how data are being collected. This is especially troublesome when we consider the context of AAL technologies: older adults are a vulnerable population that traditionally do not have advanced technological knowledge [16,23,30,31].

Older populations using AAL technology are at a high risk of being subjected to security and privacy violations because of the mishandling of their data, in which mishandling is characterized by any use of data that are different from what was consented to by the user. Older adults (and technology users in general) must know what they are consenting to and how to manage their consent at any time. In addition, many technologies only allow a *binary* consent in which users consent to all data or no data being collected. Users do not have any choice of which health variables they can give or revoke consent to.

Blockchain

Data ownership, security, anonymity, and privacy are complex topics, and, as exemplified above, the challenges of obtaining consent for increasingly advanced methods of data collection, use, and disclosure call for new solutions to imperfect consent procedures to protect the safety of individuals. One possible solution is using blockchain.

This technology can be seen as a distributed ledger formed of data structures known as blocks, equipped with cryptography techniques to enable trust among parties while being operated by a peer-to-peer network of computers [1,32]. Each computer forms an independent node on the network and maintains a copy of the ledger, which is regularly updated to ensure that it remains consistent with the other copies.

A blockchain can be private or public, depending on its intended application. A public blockchain is open to anyone who wishes to browse its contents or participate in the network [33]. Public blockchains are the most well-known blockchain applications because most cryptocurrencies take advantage of their features to enable trusted monetary transactions without the need of a trusted third party. In contrast, a permissioned or private blockchain allows only authorized users to browse its contents or participate in the network [34]. This type of blockchain solution is appropriate for sensitive or highly regulated information management environments, such as health care data.

A blockchain network receives transactions when two or more users want to transact information between them. Transactions from users are broadcasted to the network, validated, and grouped into a block by network nodes known as *miners* [35,36]. The transactions of a blockchain are data structures modeled to represent real-world processes and objects. For example, a transaction can be modeled to represent monetary transactions [37]; transfer of the ownership of a car [38]; the current state of a business object [39]; or, in our framework, the current state of informed consent of a patient.

Transactions sent to the ledger are secure and private without the necessity of a trusted third party because users of a blockchain do not use personally identifiable information as credentials when sending transactions to the ledger [1,32]. They use techniques such as cryptographic proof of ownership, in which every user has a private and public key pair for each transaction they submit to the ledger [40].

The sender signs the transaction with the private key, and a unique signature is generated and sent, along with the information of the transaction, to the public key of the receiver in the network. Every new blockchain transaction is broadcasted to all miners of the network who concurrently verify the transaction for proof of the private key ownership of the sender and verify whether the contents of the transaction are valid. For example, in the blockchain of Bitcoin [35], all transactions are checked for the private key ownership of the sender to ensure that the sender has enough funds (bitcoins) to send the transaction.

Miners utilize the public key of the sender to verify whether the signature of a transaction is valid. If the signature is not valid, it indicates that the original signature of the sender is wrong or tampered with, resulting in the network rejecting the transaction. The proof-of-ownership method prevents transactions from being corrupted or tampered before being added to the blockchain [34].

After transactions are validated, miners group them in a block; however, before adding it to the blockchain, they must complete a task known as mining [37]. Mining is a process in which miners compete to create a unique hash string for a new block.

A newly generated block hash contains in its composition the unique hash of the most recently added blockchain. The linkage of blocks' hash creates a cryptographical heritage that enables blockchain to tamper-proof its information history. For example, if an attacker wants to change the transaction information contained in block number 50, and miners from the network are currently working on block number 100, the attacker must generate the unique hash of all blocks that come after block number 50 until they reach the end of the chain and generate the unique hash of block number 100 before all the other miners of the network finish working on it. For this type of attack to be successful, a significant amount of processing power from a single miner is required. More precisely, a single miner would have to produce more processing power than the entire network to achieve success. Hence, these types of attacks are improbable [41,42].

In typical cryptocurrency blockchain solutions, such as Bitcoin, miners compete to finish mining a block because the winner gets rewarded for completing the task. This process is known as proof-of-work [35], and one of its major disadvantages is the vast amount of electrical power needed to keep networks running. Newer blockchain solutions, such as Ethereum, take advantage of modern validation processes such as proof-of-authority. In this process, instead of wasting computational power to validate and hash a block, credentials or any other relevant fact about the miner is used to accept a new block as valid or not.

Proof-of-work and proof-of-authority are used by blockchain technologies to help the network achieve consensus. In all consensus methods, a consensus is achieved by the ledger when the majority of nodes in the network agree that the block is valid and add it to their local copies. The consensus process starts when the first miner in the network finishes mining a block. It first adds the new block to its copy of the ledger, then follows to broadcast the new block to its neighboring nodes. Each node that received this new block verifies the unique block hash for validity, and if it is valid, it adds the block to its copy of the ledger. The node then follows to broadcast the new block in the same manner as the miner. This process is repeated until part or all of the network agrees with the current block. Some blockchain solutions define that when more than 50% of the network agrees with a block, consensus is achieved. There is a plethora of blockchain solutions that offer different types of consensus mechanisms. As for our framework detailed in the next sections, we used hyperledger fabric (HF) [38] to take advantage of the pluggable consensus feature, which allows for future changes in the consensus model. This feature is essential to our framework because we cannot instantiate an application that is not capable of adapting to comply with future regulations. More details on the consensus mechanism of HF are presented in the *Results* section.

This paper explores the creation of a blockchain platform for consent management in health care, specifically in the context of AAL. We discuss a general methodology for identifying blockchain use cases developed by Gorenflo et al [43], and we apply this methodology to consent management for AAL technologies. With the identification of consent management as a prime use case for blockchain, we expand upon previous work done by Bublitz et al [1], deriving from their general software architecture for surveillance activities, a conceptual framework for blockchain in AAL consent management. To develop this proposed conceptual framework, we researched which blockchain technology was the best fit. In the following sections, we discuss the methodology for identifying trust issues and software architecture, which served as a basis for our work. Next, we present our framework and explain how it differs from related work.

Methods

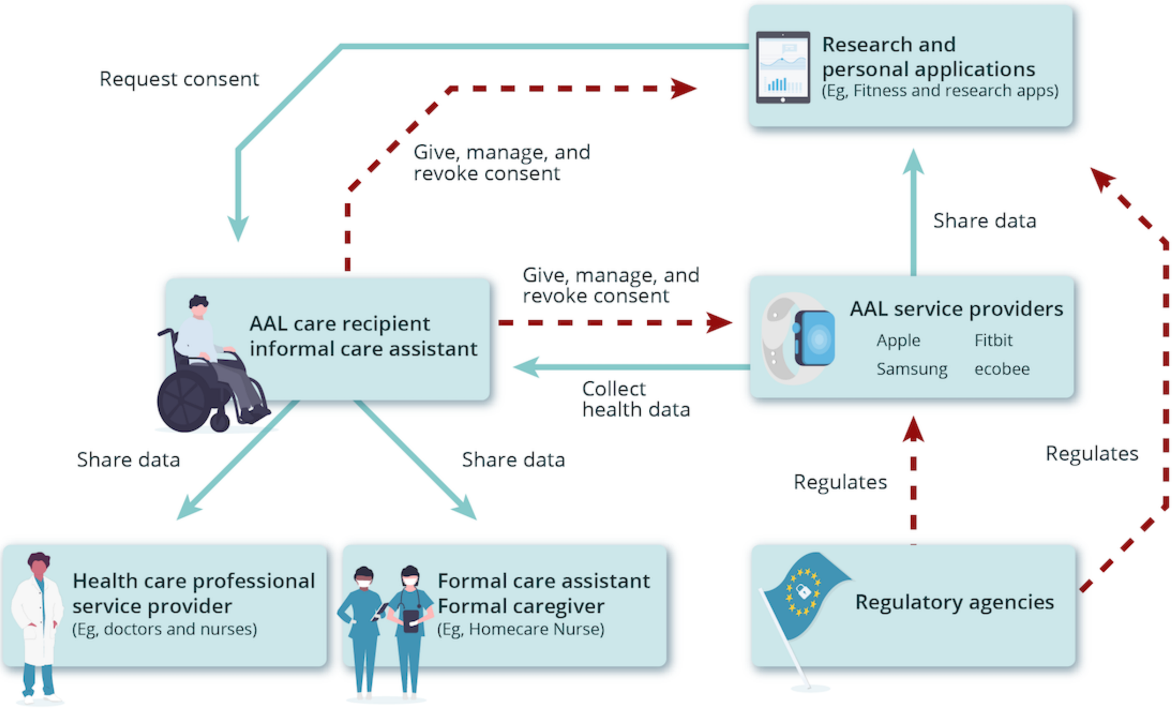
Mapping of Trust Issues

Gorenflo et al [43] defined a general methodology for identifying use cases of blockchain based on the identification of trust relations. This robust approach involves the following sequential steps relevant to this paper:

1. Identify the parties and trust relations between them. If a relationship does not have the required level of trust necessary to achieve the goal of the relationship, it should be marked as a trust issue.
2. Design a minimal blockchain system that resolves the trust issues.
3. Migrate the rest of the existing system to the new blockchain system if such a system exists.

In this study, we followed this methodology to model the consent management process for data collection in the AAL technology space, in consultation with AAL experts, resulting in the diagram presented in Figure 1.

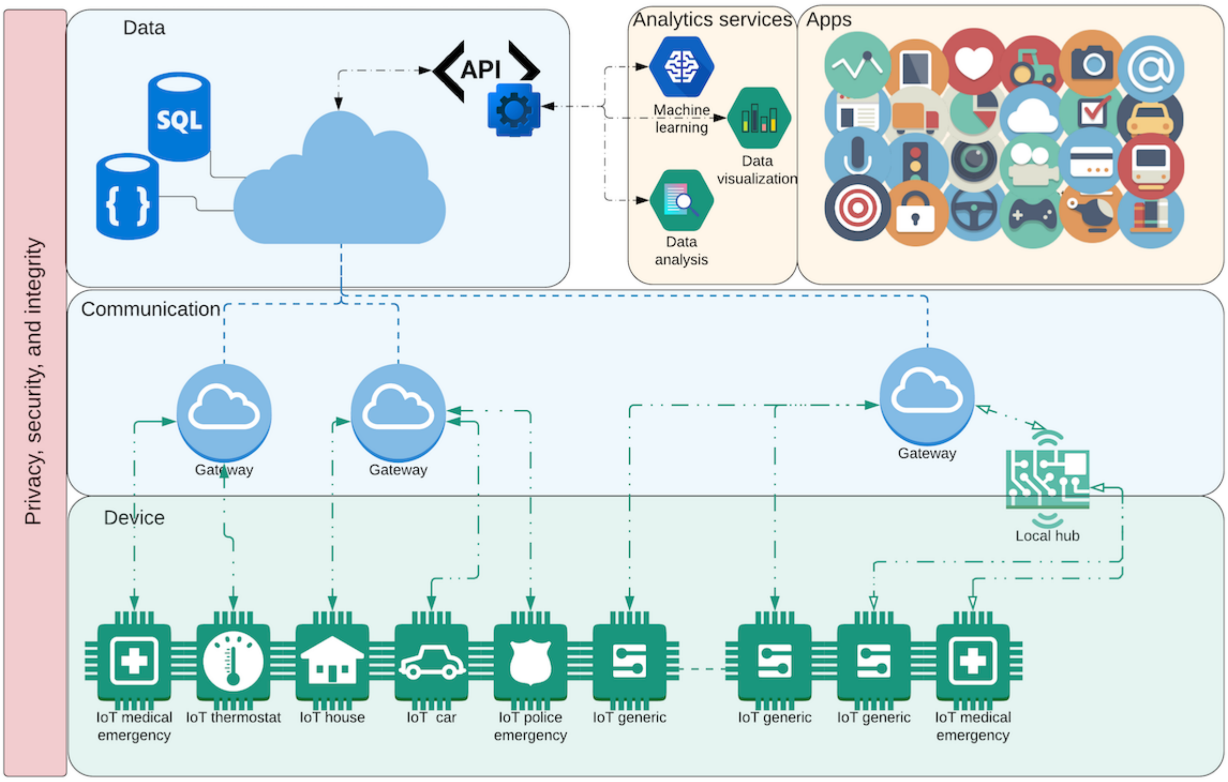
Figure 1. Consent management process and trust issues in active assisted living. AAL: active assisted living.



Framework Development

Once the trust issues were identified, we began work on a conceptual framework that focused on blockchain to mitigate these issues. This framework is based on a general framework architecture for monitoring and surveillance activities created by Bublitz et al [1] to address most of the requirements for the creation and use of IoT systems. The main layers of the architecture, as exemplified in Figure 2 [1], are device, network, data, applications and privacy, security, and integrity.

Figure 2. General architecture for the pan-Canadian surveillance system. API: application programming interface; IoT: internet of things; SQL: structured query language.



Results

Trust Diagram for Consent Management

The trust diagram in Figure 1 follows the nomenclature from the International Electrotechnical Commission or Systems Committees Active Assisted Living [1,17,44]. The identified trust issues are highlighted in red on the arrows included in the trust diagram.

An AAL Care Recipient (or an Informal Care Assistant, a nonprofessional caregiver) uses AAL technology. These users need to manage their consent to data collection and use by the manufacturers or owners of the technology, the AAL Service Providers, and any third-party application that collects the data gathered from the technology, represented as Research or Personal Applications. Trust issues arise between the AAL Care Recipient and entities consisted of Research or Personal Applications and AAL Service Providers because these entities may use the collected data in different ways than what was consented to. Another trust issue between these entities and regulatory acts, such as the Personal Information Protection and Electronic Documents Act (PIPEDA; in Canada), Health Insurance Portability and Accountability Act (HIPAA; in the United States), and General Data Protection Regulation (GDPR; in the European Union), which regulate the collection of personal data, is that the entities may be illegally using the data regarding the current legislation, thus violating the rights of the users.

The diagram clearly demonstrates relations without the necessary level of trust in the consent management process. Blockchain technology can provide an immutable and timestamped log of consent, making the process more transparent for everyone involved. This is extremely important as consent management, traditionally, is subject to many shortcomings and misconducts [45], and not following correct consent procedures can have tragic consequences. For example, in 2016, the trial testing of the drug BIA 10-2474 in France caused the death of a participant and hospitalization of 5 others, an event classified as that of “exceptional gravity, unprecedented in our country” by the Minister of Solidarity and Health [46]. Preliminary investigations showed that after major neurological effects were discovered in 1 patient, the researchers did not obtain reconsent from other participants, allowing them to continue in the trial despite clear dangers to their safety [46]. If the researchers used a platform such as the one proposed here, all their interaction with the participants regarding consent management would be recorded with a timestamp, making it extremely difficult for the researchers to not obtain reconsent as this misconduct would be easily auditable.

With a blockchain platform for consent management, users are able to monitor and manage their consent in real time and with granular variable control, for example, by giving informed consent for certain types of data to be collected but not others or revoking their consent at any time. This is in alignment with privacy regulations, such as the data protection by design of GDPR and right to data deletion by users at any time [47]. This information will be immutably stored in the blockchain and may be accessed by all entities with permission to do so. Users will also be able to ensure that the forms were approved and cryptographically signed by the review ethics boards, ensuring that the researchers are not withholding any information.

Data collectors will also benefit as measures taken to ensure ethical and legal requirements throughout the data collection process will be clearly documented and auditable.

A feature of blockchain called smart contracts [1] can also be leveraged to improve the consent management process. Smart contracts can be seen as codified contract agreements, and because blockchain is an immutable ledger, terms of a contract written into software and embedded in the blockchain will always guarantee the fulfillment of these terms [1]. Smart contracts can be used, for example, to bind SDMs to an individual or to check whether the consent process is compliant with different privacy regulations across territories. Ultimately, with blockchain technology, it will be easier to obtain, track, and update informed consent.

The Blockchain: HF

The HF [48] platform was selected for use as the underlying blockchain technology as it provides the tools to achieve the goals of our proposed conceptual framework (Textbox 1). HF is an open source, permissioned, distributed ledger technology (DLT) platform [38,49] developed to be used in the contexts of enterprise. All enterprise interactions are performed in a private network environment called hyperledger fabric network (HFN). An HFN provides blockchain services so that client applications and network administrators can interact with the HFN, proposing new transactions and querying the blockchain. In an HFN context, a consortium [50] comprises 2 or more organizations on the HFN that need to transact business with each other.

In HF, an organization [51] is a logical driven group of members that can represent big or small corporations. Each HFN organization is composed of members called peers [52]. Peers are responsible for hosting ledgers, smart contracts, handling changes to the ledger, communications between different organizations (channels), and managing external requests from applications. An endorsing peer is responsible for executing smart contracts over a transaction proposal from a client to verify its validity before it can be added to the ledger. A committing peer is a member who keeps a copy of the ledger without any smart contracts, meaning that this is done to keep multiple copies of the ledger to avoid single-point failures. Ordering peers and organizations [53] are responsible for keeping the communication sound between the participants of a channel; however, those components are beyond the scope of this paper.

Organizations can communicate with each other on an HFN by creating and using communication channels. Communication channels are private communication environments for all or a subset of the organization’s members of a consortium. New organizations are enrolled in the HFN through a trusted membership service provider (MSP) [54], which is responsible for issuing and validating certificates and user authentication.

The conduct of business transactions between organizations of an HFN consortium is stored in a ledger. A ledger stores both the current state and the history of states from a business object. To keep a business object, the current state HF uses a database called the world state [55], which is explained later in this paper. To store the history of the transactions of a business object, HF utilizes a blockchain data structure [38]. Together, the blockchain structure and world state are what allow the ledger to hold an immutable history of states of business objects (blockchain) and provide fast access to the most recent state of a single business object.

Participants of the HFN achieve consensus by following a transaction flow that takes into account the endorsement policy assigned to smart contracts [56]. This means that transactions can only be accepted into the ledger if all, or a subset of, the endorsing peers of a consortium approve the transaction. Smart contracts in HFN are a software representation of a contract that governs the processes that alter the state of a business object. In HF, smart contracts are packaged into a structure called chaincode, which consists of multiple smart contracts. Each chaincode is deployed into an HFN channel with an endorsement policy assigned to it. The endorsement represents which organizations must sign a transaction so that it is accepted and added to the ledger.

This transaction flow from HF relies on identity validation for members of the network. All transactions are first sent to the network as a proposal that needs to be endorsed by all HFN organizations included in the policy of the smart contract. The proposal is signed using the cryptographic credentials of the user that generate a unique signature.

Endorsement organizations from HFN verify transactions for several factors: (1) a well-formed transaction proposal, (2) whether the proposal is not repeated, (3) checks with the MSP [54] to verify whether the proposal signature is correct, and (4) then the MSP verifies whether the identity of the proposal submitter has the authorization to perform operations in that channel.

Input parameters are fed to the chaincode that is going to be executed in the current world state. The result is returned after chaincode execution alongside the peer signature to the client as a proposal response. Finally, the client inspects the response from peers, and if all the necessary signatures are valid, the transaction is created.

The client application broadcasts transactions to the ordering service. The service, in turn, creates orderly blocks of transactions and sends them to the channel’s peers. Each peer verifies the block transactions and signatures before adding the block to the ledger. HF is not the only appropriate platform that can provide the necessary infrastructure to instantiate our conceptual framework. For example, Ethereum [57] is a blockchain solution that provides full support for distributed applications over a blockchain network. Even if the main network of Ethereum is public, which is not ideal for the health care domain, it can be instantiated into a private network environment, thus becoming compliant with the privacy and access control requirements of the present conceptual framework. We refer to Modum [58] as an example of an implementation of a supply chain management system applied to Ethereum. Modum enables pharmaceutical companies to monitor the temperature and humidity of medical products by using smart contracts to verify sensor data during their life cycle.

Other honorable mentions of blockchain platforms that are appropriate to this framework instantiation are the IBM blockchain platform [59], the Multichain Private blockchain [60], the Hydra chain [61], and the BigchainDB [62]. It is also worth mentioning that using these technologies will require some changes in the HFN [63] component of the framework.

Textbox 1 contains the minimum requirements of the solution mapped over the attributes of HF, providing a checklist for other blockchain platforms to instantiate the framework.

Textbox 1. Active assisted living conceptual framework requirements and features of hyperledger fabric that support them.

Privacy of patients' information

- Hyperledger fabric (HF) uses cryptographic material to control access to the ledger. Applications from authorized organizations keep information from patients in their infrastructure. HF only stores anonymized consent information from unique keys created for each patient.

Private communications between stakeholders of the informed consent process

- HF maps stakeholders from a network into organizations. Each organization can be a part of a channel composed of some other organizations that constitute the consortium. Communications inside a channel are private to their members.

Scalable and fast querying of the ledger

- HF stores the most recent state of a business object in a state database called world state. Every time an informed consent state changes or a new informed consent is created, a new transaction is inserted into the ledger. However, the world state only updates the business object if it already exists, otherwise it is created. These characteristics also make HF more scalable as querying the complete history of the ledger is not required for determining the most recent informed consent status.

No cryptocurrency

- HF has no cryptocurrency, which complies with our solution's intention to provide the simplest solution possible to our end users, people using active assisted living.

Security of information

- As HF is a private and permissioned blockchain platform, the security of information is much less susceptible to a privacy breach.

Immutability of ledger history

- HF provides an immutable ledger natively.

Accountability of informed consent transactions

- With HF, smart contracts can be created enforcing standards. These smart contracts can be configured to explicitly require the endorsement of an organization that represents an auditing authority.

Software development kit (SDK) for development of end user applications capable of interacting with the ledger

- For our conceptual framework to be instantiated, the creation of web applications capable of interacting with the blockchain network to create and retrieve consent transactions is necessary. HF provides such SDKs in NodeJS, Java, and Go languages.

Smart contracts capabilities

- Smart contracts are needed to ensure the enforcement of standards and endorsements that must be respected so that a new informed consent transaction is accepted and stored in the ledger.

Custom network policies

- Such policies are needed to control network interactions between organizations. The informed consent process has its own unique policing needs, such as endorsement policies, hence the need for customization.

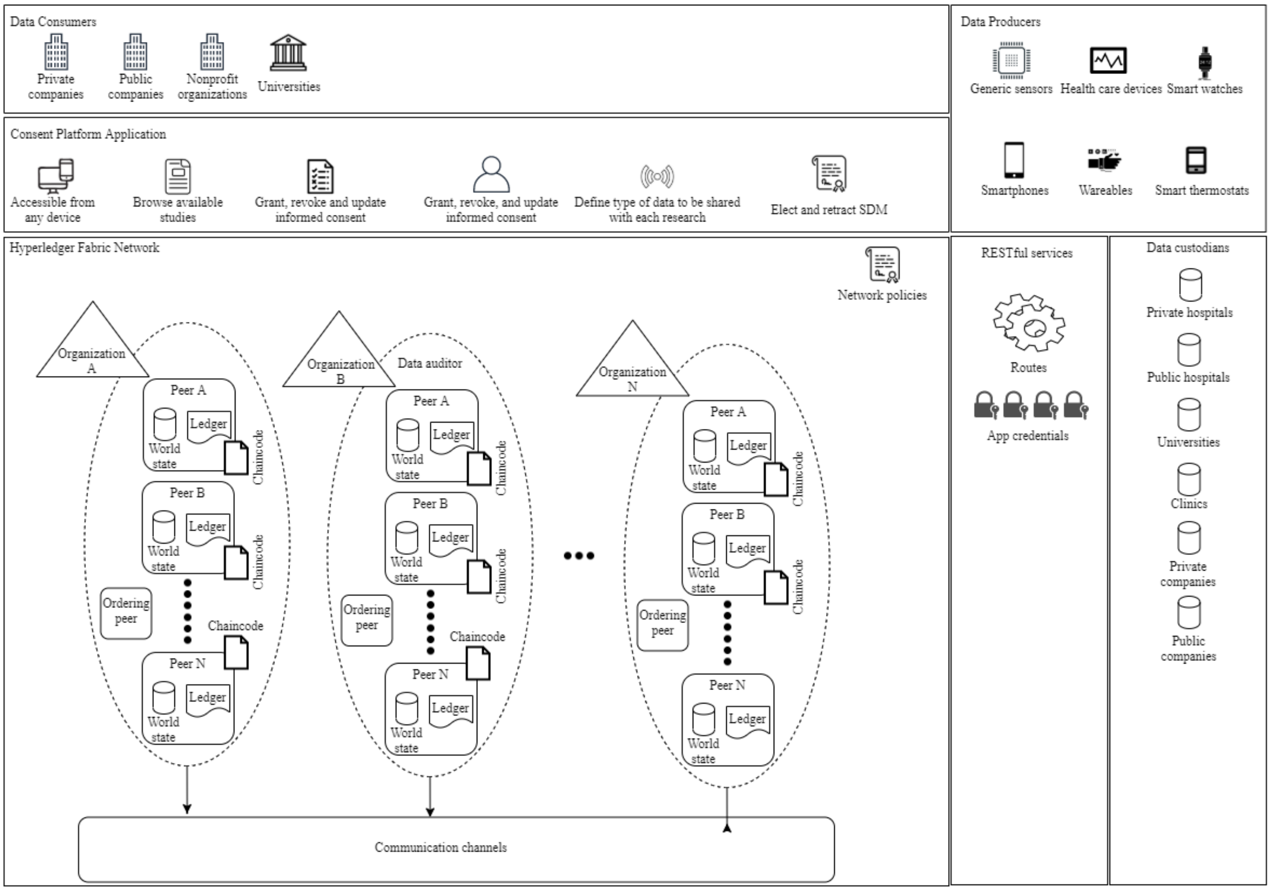
Extensibility of the network

- For our solution's relevancy, the capability of adding new organizations, policies, channels, and smart contracts without having to start a new network is imperative. HF provides the plug-and-play capability for adding new entities into an existing network.

Conceptual Framework for a Blockchain Consent Platform in AAL

As previously mentioned, the main objective of the conceptual framework is to allow a plethora of smart *devices* data to be used by different health care stakeholders while providing trust, security, and privacy. Figure 3 illustrates the conceptual framework components, one for each domain of services and health care data. All stakeholders are either mapped into an HF organization or are end users that interact with the framework through the web application of the consent management platform. Each square in Figure 3 represents a different framework component, and each component maps services, applications, participants, and infrastructures that need to interact with the platform to complete the instantiation of the framework.

Figure 3. Consent management platform conceptual framework. SDM: substitute decision maker.



For example, the data consumer component interacts with the consent management platform component to create requests for informed consent from AAL patients. The data producer component must interact with the consent management platform component to grant informed consent to data consumers and to register the consent into the ledger. The consent management platform component interacts with the RESTful (representational state transfer) services component to send transactions to the ledger. The RESTful services component acts as the middleware between the platform front-end applications and the HFN component.

Finally, the data custodian component interacts with the consent management platform to ensure that a data consumer has the proper informed consent registered in the blockchain before sharing patient data.

The *SecurityModel* section discusses and explains the platform security aspects in details (Figure 3), along with what features from HF were utilized to support the goals of the framework.

Security Model

As seen in Figure 3, the conceptual framework is built on top of an HFN; hence, all communications between organizations and their clients are secure by state-of-the-art authentication methods such as x509 certificates and asymmetric key pairs [38]. We derived our security model from the architecture of HF because it offers cryptographic certificates and keys necessary to access the ledger.

The consent management platform application utilizes HF ordering service [53] to access the HFN. HF applications are bound to an organization, and each time a new transaction is sent from an application, the network verifies the credentials of the organization and users before accepting transactions. If the credentials of the HF application are valid, then the transaction can be submitted to the ledger.

All transactions sent to the ledger are previously checked by cryptographic access control guards, which ensures that only authorized users can access the HFN. Chaincode is used to ensure that all transactions are signed and verified by the necessary endorsing organizations. For each organization endorsement, digital signatures are used to provide data authenticity before adding the transaction into the ledger. If an endorser organization rejects the authenticity of an application's user, then the transaction will not be inserted into the ledger.

Cryptographic authentication mechanisms are simplified by a feature from HF called MSP [54]. An MSP abstracts complexity involved in issuing, validating, and certificating users' and organizations' identities. An MSP is configured to have unique identities and rules that govern these identities. The HFN utilizes the identities of users to control access to the ledger. For example, a user with a *client* type of identity can only transact on the network, but an identity of type *peer* is allowed to endorse or commit transactions to the ledger.

Data Consumer Component

The data consumer component represents stakeholders that are interested in consuming data for a variety of different reasons, such as remote patient monitoring (RPM) [64-67]. To access health care data, obtaining informed consent from participants is essential to protect all participants, and ensuring safe and ethical procedures are in place is also essential [18,22,24,45]. The framework provides such interactions through data consumers interacting with the HFN component. It is also essential to explain that data collected for a data consumer can be real time or historical. This means that levels of security and awareness requirements from participants can differ, thus forcing the utilization of stricter network policies and smart contracts based on the characteristics of the data.

Data Producer Component

The data producer component is constituted by stakeholders that provide health care data in the realm of the conceptual framework [37,68]. Collected data can be from passive or active monitoring [69]. The former relates to data that are continuously or periodically sent to the cloud (eg, smart thermostats), and the latter is the type of data that requires actions to be taken on the part of data producers (eg, clinical exam) [69]. Any request for active data use from a data consumer must be consented to by a data producer for a data consumer to gain access to it. Consent is given when a data producer interacts with the consent management platform and explicitly gives consent to the data consumer. The consent management platform is responsible for creating the transaction and sending it to the HFN. If the transaction is valid and endorsed by all the appropriate peers, then the transaction is stored into the ledger and the world state is updated.

Data Custodians Component

The data custodian component contains stakeholders that are responsible for stored health care data [70,71]. Aside from the PIPEDA and HIPAA considerations with regard to security and privacy standards that all data custodians must adhere to, they must only provide access to data by another stakeholder from the informed consent process if the informed consent exists and if it has not expired. It is worth mentioning that a data consumer may also be a data custodian in some cases, but informed consent must be obtained in the same manner as previously noted. If a stakeholder is a data custodian, in the holistic view of the framework, this does not mean they are entitled to use health care data freely [69]. Data custodians are represented as HFN organizations in our framework, so that they can be part of the channels responsible for storing consent information. Therefore, by querying the world state, they can verify whether the consent given to a data consumer is valid.

HFN Component

The HFN component represents the mapping of all the stakeholders of the conceptual framework into HFN organizations. By creating this mapping, it is possible to create consortiums capable of having private communication channels with their ledgers and smart contracts. These channels are used by the organizations to query, create, and update informed consent state transactions that are stored within the ledger. New organizations can be added to the HFN at any point; thus, the organizations in the framework architecture are generalized. The ledger contains historical data from all the transactions between the organizations, which allows for data auditing from any of the network participants, consequently ensuring that trust evolves as a result of design.

Another feature that comes by design, thanks to the use of HF, is the world state. The world state possesses the current values of a business object (in our case, informed consent data). This is necessary for our conceptual framework, as it would be expensive and time-consuming to go through all transactions stored in the ledger to find the most recent state of a consent document. In HF, the world state is a NoSQL (No Structured Query Language) database that stores key-value pairs as identifiers of each state. The key-value pairs can be, for example, the pair patientID-researchID. Every time a new transaction from this key is stored into the ledger, the key-value or pair state stored in the CouchDB will also be updated.

Finally, as per the current version of HF, we utilize the consensus mechanism from the Raft protocol. The Raft protocol is based on a leader and follower model in which a node is elected as leader and the rest as followers. The Raft protocol is also a crash fault tolerance service that complies with our requirements of ensuring access to patients' consent information at all times [72].

Data Auditor Component

A data auditor can be any organization that represents regulatory authorities such as the PIPEDA and HIPAA. Having data auditors present on the network is important because they ensure that law and regulations are enforced in the HFN by using smart contracts and endorsement policies. By being a participant of the HFN and possessing the endorsement of peers included in the proper channels of interest, for example, channels that store information of the data auditor's jurisdiction, a data auditor can audit the information of multiple ledgers if needed. This HF capability allows the process of informed consent management to be auditable by different authorities, granting the conceptual framework the capability to adapt to new rules and policies that new data auditor added to the HFN might bring to the consortium.

Another possibility is that a data auditor, through the creation of policies, can force transactions to satisfy a set of requirements through smart contracts and network policies. Smart contracts enable the enforcement of standards for data privacy and security that prevent breaches of privacy and misuse of information. Through the proactiveness of a data auditor, it is possible to enhance the informed consent process as it can continuously verify the validity of consent. A data auditor can act upon expired or invalid consent information by being a member or founder of the channels in which the organizations under its jurisdiction are present.

Consent Management Platform Component

The consent management platform is a web-based platform that groups all the functionalities necessary for the processing of granting, revoking, and managing consent. First, to prevent a lack of access from stakeholders because of technology limitations, the framework proposes the use of responsive web applications that can be accessed by any device. The platform uses HF's NodeJS software development kit (SDK) [73] to interact with the HFN from the front end. Different endpoints can be used to control actions allowed only by certain types of users and organizations, for example, a patient (data producer) can elect an SDM to make decisions for him but cannot create requests for consent like a data consumer.

Patients can log into the platform to browse available data consumer requests and choose which to enroll in. For example, if an AAL patient wants to be monitored by an RPM center, the RPM center, as a data consumer, logs into the platform and creates the request for consent to be accepted by its patients. The AAL patient can grant consent to the RPM center through the platform's web application. In doing so, a transaction is created and proposed to the HFN and, if accepted, is stored in the ledger. This ensures that all the stakeholders from the transaction now possess a copy from the current state of consent, and the data consumer can start collecting the patient data from data custodians.

The consent management platform also provides web interfaces that allow patients to revoke consent and update consent. Another feature offered is that data producers can choose to share specific types of data for each request for consent. For example, an RPM center may request 5 types of data to be shared by a patient. However, the patient may choose only to share 2 types if they desire; although this makes the expected data received by the data consumer less predictive, it empowers data producers at the same time. This process ensures accountability, security, and transparency for the informed consent process, helping to improve traditional methods of obtaining consent.

The consent management platform can also be used to enhance the process of electing SDMs for patients. A patient can request that a user becomes their SDM. If the appointed SDM accepts the request, the information about the SDM elected is stored into the ledger and becomes available to stakeholders. Just like an informed consent state stored into the ledger, the SDM state can also be revoked or updated based on the circumstances. Such features can help to empower people under AAL conditions to quickly elect trustworthy SDMs without having to go through current bureaucracies that govern the substitute decision-making process [74].

Finally, the representational state transfer services layer is responsible for managing appropriate access to the consent management platform and for keeping meta-information about the users' identities in the ledger. If any other functionalities are needed, the conceptual framework can be extended to encompass new functionalities and components.

Chaincode

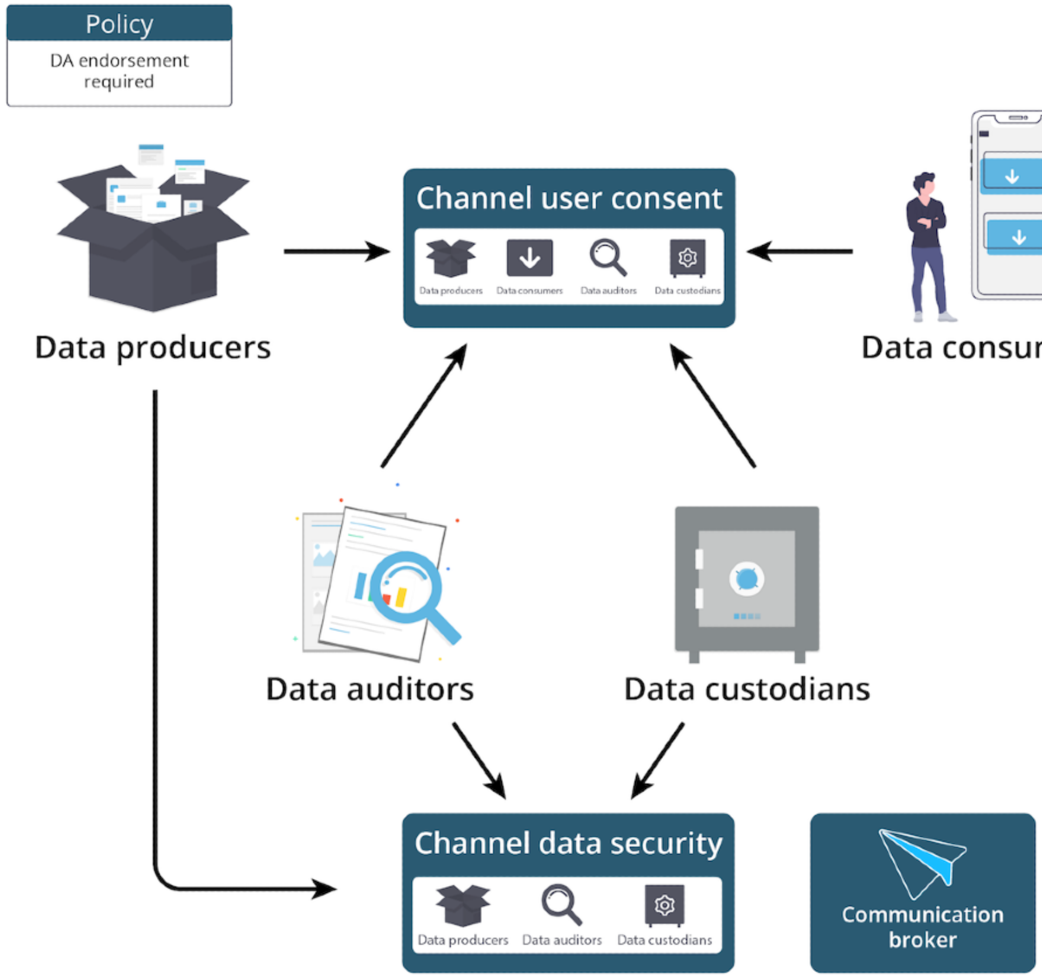
The utilization of chaincode (smart contracts) by the framework ensures that only registered applications can send a transaction to the ledger and that the network's required organizations endorse only valid transactions. A set of endorsement policies is defined for each new chaincode deployed into the network. As we are proposing a conceptual framework, at this stage of our research, we still do not have abstract contracts that can be extended by other developers. However, in the *Framework Instantiation* section, we explain the development of our prototype's chaincode.

Our first implementation of the chaincode will serve as a subject to further enhancements for the framework. In future work, we will study all contracts and their commonalities to create abstract contracts that can be instantiated by developers who want to adapt the framework for other domains.

Framework Instantiation

To exemplify the instantiation of the framework using HF, Figure 4 shows an abstraction of the developed HFN. The consortium is composed of 4 organizations that represent minimum requirements to allow all stakeholders in the system to perform the required tasks.

Figure 4. Organizations instantiated in the framework. DA: data auditors.



In our network, the 4 organizations that compose the consortium are data producer, data auditor, data consumer, and one data custodian. The members of the channel, channel user consent, have access to the ledger containing informed consent information. Before any new consent is given and stored into the blockchain, the data auditor member of the channel must agree and endorse the transaction containing the informed consent (as stated by the network policy in Figure 4). The data custodian, as a member of this channel, knows that the data auditor agreed to that consent if the transaction is valid and stored into the ledger. With the transaction validated, the data custodian is authorized to give access to its health care data to the data consumer as long as the data custodian respects the restrictions defined by a patient or SDM.

The data auditor organization is also a member of a different channel of communication called channel data security. This channel will hold the members of the network responsible for ensuring that the data custodian has the proper consent from users to store their information on their servers. Finally, the communication broker represents an organization responsible for routing the communication between the end user and the communication channels.

We used NodeJS SDK of HF to implement the chaincode responsible for verifying, adding, and querying informed consent of the patients and SDM states for members of the channel *User Consent*. The chaincode of the channel is defined by 2 contracts: contract one (C1) and contract two (C2). C1 is used to manage the informed consent state and to manage substitute decision maker states. C1 ensures that all fields of the informed consent transaction are present and valid. C1 checks for patient and research identification fields; the number of sensors; and, if for each type of sensor shared, the periods of consent are valid. C2 is responsible for managing SDM states from members of the *User Consent* channel. C2 checks for SDM and patient identification, start and end dates of the SDM validity, and the current status of the SDM-patient relationship. A code snippet implementation using NodeJS SDK of HF for C2 is presented below (Textbox 2).

If all requirements are successful, then the organizations, after executing the chaincode, endorse the transaction and return the proposal result to the sender. After finishing the proposal process, the new informed consent status is added to the ledger.

Textbox 2. Code snippet from the chaincode to add a new substitute decision maker.

```
/**
@function addSDMState
@param {*} ctx
@param {*} patientID
@param {*} SDMID
@param {*} rlStatus
@param {*} startDate
@param {*} endDate
*/
async addSDMState(ctx, patientID, SDMID, rlStatus, startDate, endDate) {
try {
// Instantiate a new SDM state to be added to the ledger
let substituteDecisionMakerState = SubstituteDecisionMakerState.createInstance(patientID, SDMID, rlStatus, startDate, endDate);
if (substituteDecisionMakerState) {
// add to the ledger
const response = await ctx.substituteDecisionMakerStateList.addSDM(substituteDecisionMakerState);
// return ledger response
return response;
} else {
console.log(“Error at addSDMState”);
return null;
}
}catch (error) {
console.log(“Error at addSDMState”);
}
}
```

Discussion

Enhancing Consent Management

Ensuring proper informed consent is a major concern for data collection and use [22,45]. Our proposed solution facilitates health data sharing by different stakeholders while increasing transparency and trust. The biggest advantage of our solution is that it was developed based on a systematic process of identifying and mitigating the trust issues in the consent management process [43]. The framework minimizes all trust issues indicated in Figure 1 by providing a virtual space to manage consent, powered by blockchain to provide an immutable and timestamped log of user consent for data owners and collectors. Consequently, AAL care recipients and informal care assistants will have a much better understanding and control over what data they are sharing, with whom, for what purpose, in what manner, and for what time period. Furthermore, it will be easier for regulatory agencies to audit if AAL service providers or research or personal applications are using data for purposes other than what was originally consented to, as they will be able to access the log of consent of the blockchain.

The US Food and Drug Administration (FDA) reported in an overview of clinical trial inspections from 1977 to 2009 that 28% of trials had deficiencies related to inadequate consent forms [45]. In 2012, the FDA cited the main deficiencies related to consent, among others: the failure to obtain informed consent; use of expired, incomplete, or nonvalidated forms; failure to provide copies of the forms to study subjects; missing documents; and changes made to documents by hand and without the approval of ethics review boards [22]. Our platform will provide a secure and immutable virtual space where all stakeholders—data owners, consumers, and regulators—will have complete transparency and surety of the entire process, making the process safer for all involved and ensuring that cases such as the trial of BIA 10-2474 will not happen again.

In short, the proposed conceptual framework tries to generalize the process of consent management for all stakeholders in the health care domain, allowing a clearer understanding of possible interactions and functionalities of the consent management process and, ultimately, providing more transparency. As presented by Novitsky et al [18], this was an important characteristic to help address inefficiencies in the process of consent management, especially for vulnerable populations.

The framework also aims to take full advantage of the features of HF to allow for robust and complex control over the process of obtaining informed consent. Data auditors are an example of such goals, in which they can remove bad actor organizations from the network if they do not comply with the rules.

Related Work

Several companies explore consent management and blockchain. However, they differ significantly from our solution.

Hu-manity.co, for example, developed a mobile app with IBM blockchain to help individuals manage consent for the use of their personal and health information. This app gives users a title for their digital data, declaring them as the user's property [75]. Unlike our proposed platform, this solution does not store the user consent information on the blockchain.

Another company, Bitfury, is producing a blockchain-based consent management system for research and medical data. This solution is closer to what we are envisioning as a blockchain platform [76,77]. The same is true for solutions in academia, such as the work of Benchoufi and Ravaud [26] that use blockchain to provide a timestamped log of consent for clinical trials. We differentiate from these works not only by focusing on IoT and AAL, and outside clinical and research contexts, but also by providing a granular variable control for users, allowing them to manage consent for different data types and periods (eg, user 1 gives consent for the temperature to be collected but not their movement, between June and July).

An interesting related solution is MedRec, developed by researchers at the Massachusetts Institute of Technology. Unlike previous works, this solution was not created to improve the consent management process but to minimize interoperability issues between electronic health records (EHRs) while providing a transparent view of a patient's full medical history. MedRec uses smart contracts to encode metadata with references to medical data from multiple medical data sources, which includes information regarding ownership of the data. Patients permit data to be accessed and shared [25]. In MedRec, consent management is not an explicit concern, but it appears as a form of access control for medical data. Although our solution considers access to data, as described in the sections below, we focus on consent management and improvement of trust relations in this process. In addition, MedRec deals with medical data stored in clinical systems and already collected, whereas our platform deals with real-time, patient-generated health data from sensors and AAL technologies deployed in the real world. Although our solution is distinct from solutions that deal with EHR interoperability, having proper informed consent is always a concern when dealing with personal data. Our proposed solution is a possible facilitator such that future health data sharing solutions do not need to develop tools to manage informed consent.

Limitations

The limitations of this work include the lack of implementation and real-world deployment of our proposed conceptual framework. Therefore, future work will focus on the development of a blockchain-enabled platform for consent management. Additional limitations include the fact that, for our architecture to be implemented at scale, the participation of federal and regional agencies is required to make up the participating nodes in the blockchain. In addition, data collectors will need to enroll in our platform and allow integration with the framework for the users to be able to give proper consent. Given that the platform will increase transparency and compliance with regulations for data collectors and owners alike, the stakeholders will benefit from engaging with such a platform, ensuring their participation. This means that the success of this platform depends upon the collaboration of several governments and industry partners interested in improving the current security and privacy issues.

Another limitation of our work is that the current version of the framework is not prepared to share health care data between data custodians and consumers. The platform serves as a tool for obtaining, managing, and consulting informed consent and SDM information. Custodians can use the platform to ensure that a data consumer has collected proper informed consent from the owners of the health care data they are requesting. For future work, we intend to extend our conceptual framework to be capable of offering data brokerage between custodians and consumers.

Conclusions

AAL technologies have the potential to completely revolutionize how older adults age, minimize risks, and increase independence [17]. However, this must be considered alongside the privacy implications of monitoring technologies [17,18]. Unfortunately, it is currently challenging for individuals to successfully manage their consent for data collection [1]. Blockchain is a novel technology that provides immutability and decentralization, allowing increased transparency across processes [1,32]. In this work, we modeled the trust issues existing in the consent management process of AAL technologies and proposed a conceptual framework based on blockchain to mitigate the identified trust issues. The proposed framework can be applied in different domains that deal with sensor data, such as drug supply chain [78-80] and environmental surveillance [1]. The instantiation of the platform is still in the early stages of development, but the first implementation of a fully functional application prototype has already been developed. The prototype allows researchers to request informed consent from AAL patients, and each new informed consent is wrapped into a transaction and sent into the HFN. Organizations validate and endorse the new transaction by using chaincode before it is added to the ledger. After the new block is added, participants of the network can query the ledger to check for informed consent validity before sharing patients' AAL data. Finally, although our prototype is currently not open source, we hope that researchers will use our framework to create their own blockchain applications.

Acknowledgments

This research was supported by the Canadian Standards Association Group and The Mathematics of Information Technology and Complex Systems.

Authors' Contributions

PV and PA wrote the manuscript and developed the framework. PM contributed to the conceptualization, design, and approach for the manuscript as well as to the interpretation of the argument made in the manuscript. All authors contributed to writing and revising the manuscript. All authors provided final approval of the manuscript. All authors agree to be accountable for the manuscript.

Conflicts of Interest

None declared.

References

1. Bublitz FM, Oetomo A, Sahu KS, Kuang A, Fadrique LX, Velmovitsky PE, et al. Disruptive technologies for environment and health research: an overview of artificial intelligence, blockchain, and internet of things. Int J Environ Res Public Health 2019 Oct 11;16(20):3847. [doi: 10.3390/ijerph16203847] [Medline: 31614632]
2. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
3. Wei J. How wearables intersect with the cloud and the internet of things : considerations for the developers of wearables. IEEE Consumer Electron Mag 2014 Jul;3(3):53-56. [doi: 10.1109/mce.2014.2317895]
4. Mohun J. Design for health. Nurs Times 1989 May 24;85(21):16-17. [Medline: 2734171]
5. Life in the Fast Lane: How are Canadians managing? Satistics Canada. 2016. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/171114/dq171114a-eng.htm [accessed 2020-05-31]
6. Medeiros H. [57% of the Brazilian population uses smartphones, says study]. Exame. 2016. URL: https://exame.com/tecnologia/57-da-populacao-brasileira-usa-smartphone-diz-estudo/ [accessed 2018-10-12]
7. Smartphone penetration in Argentina. Statista. 2020. URL: https://www.statista.com/statistics/621034/smartphone-user-penetration-in-argentina/ [accessed 2020-05-31]
8. Silver L, Smith A, Jonhson C, Jiang J, Anderson M, Rainie L. Mobile Connectivity in Emerging Economies. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/03/07/mobile-connectivity-in-emerging-economies/ [accessed 2020-05-31]
9. Global smartwatch unit sales 2014-2018. Statista. 2018. URL: https://www.statista.com/statistics/538237/global-smartwatch-unit-sales [accessed 2018-10-12]
10. Wearables - Canada. Statista. 2020. URL: https://www.statista.com/outlook/319/108/wearables/canada [accessed 2018-12-13]
11. de Arriba-Pérez F, Caeiro-Rodríguez M, Santos-Gago J. Collection and processing of data from wrist wearable devices in heterogeneous and multiple-user scenarios. Sensors (Basel) 2016 Sep 21;16(9):1538 [FREE Full text] [doi: 10.3390/s16091538] [Medline: 27657081]
12. Hovsepian K, al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: towards a gold standard for continuous stress assessment in the mobile environment. Proc ACM Int Conf Ubiquitous Comput 2015 Sep;2015:493-504 [FREE Full text] [doi: 10.1145/2750858.2807526] [Medline: 26543926]
13. ecobee Launches ‘Donate Your Data,’ Connects Customers to Leading Energy Researchers to Address Climate Change. Ecobee. 2016. URL: https://www.ecobee.com/en-ca/newsroom/press-releases/ecobee-launches-donate-your-data/ [accessed 2020-05-31]
14. Ecobee. 2020. URL: https://www.ecobee.com/smarthome/ [accessed 2020-05-31]
15. Barrett MA, Humblet O, Hiatt RA, Adler NE. Big data and disease prevention: from quantified self to quantified communities. Big Data 2013 Sep;1(3):168-175. [doi: 10.1089/big.2013.0027] [Medline: 27442198]
16. Vijayalakshmi A, Jose DV. Internet of Things for Ambient-Assisted Living—An Overview. In: Internet of Things Use Cases for the Healthcare Industry. Cham: Springer International Publishing; 2020:221-239.
17. Fadrique LX, Rahman D, Morita PP. The Active Assisted Living Landscape in Canada. CSA Group. 2018. URL: https://www.csagroup.org/wp-content/uploads/CSA-Group-Research-AAL.pdf [accessed 2020-05-21]
18. Novitzky P, Smeaton AF, Chen C, Irving K, Jacquemard T, O’Brolcháin F, et al. A review of contemporary work on the ethics of ambient assisted living technologies for people with dementia. Sci Eng Ethics 2014 Jun 19;21(3):707-765. [doi: 10.1007/s11948-014-9552-x] [Medline: 24942810]
19. Government of Canada — Action for Seniors report. Government of Canada. 2014. URL: https://www.canada.ca/en/employment-social-development/programs/seniors-action-report.html [accessed 2019-02-10]
20. Faiola A, Papautsky EL, Isola M. Empowering the aging with mobile health: a mHealth framework for supporting sustainable healthy lifestyle behavior. Curr Probl Cardiol 2019 Aug;44(8):232-266. [doi: 10.1016/j.cpcardiol.2018.06.003] [Medline: 30185374]

21. Important Facts about Falls | Home and Recreational Safety. CDC Injury Center.: CDC Injury Center; 2017. URL: https://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html [accessed 2020-05-15]
22. Barney R, Antisdel M. Common Problems in Informed Consent. Yale University.: Yale; 2013. URL: https://your.yale.edu/policies-procedures/other/common-problems-informed-consent [accessed 2020-05-31]
23. O'Connor Y, Rowan W, Lynch L, Heavin C. Privacy by design: informed consent and internet of things for smart health. Procedia Comput Sci 2017;113:653-658. [doi: 10.1016/j.procs.2017.08.329]
24. Gupta U. Informed consent in clinical research: revisiting few concepts and areas. Perspect Clin Res 2013 Jan;4(1):26-32 [FREE Full text] [doi: 10.4103/2229-3485.106373] [Medline: 23533976]
25. Ekblaw MC. Medrec: Blockchain for Medical Data Access, Permission Management and Trend Analysis Signature redacted Signature. Massachusetts Institute of Technology.: MIT; 2017. URL: https://www.media.mit.edu/publications/medrec-blockchain-for-medical-data-access-permission-management-and-trend-analysis/ [accessed 2020-05-31]
26. Benchoufi M, Ravaud P. Blockchain technology for improving clinical research quality. Trials. 2017 Jul 19. URL: https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-017-2035-z [accessed 2020-11-24]
27. Knifed E, Lipsman N, Mason W, Bernstein M. Patients' perception of the informed consent process for neurooncology clinical trials. Neuro Oncol 2008 Jun;10(3):348-354 [FREE Full text] [doi: 10.1215/15228517-2008-007] [Medline: 18388256]
28. Williams E. Informed consent in evaluation: informed of what, exactly? LCJ 2014 Sep 01;14:180-203. [doi: 10.18793/lcj2014.14.13]
29. Heyland DK, Cook DJ, Rocker GM, Dodek PM, Kutsogiannis DJ, Peters S, et al. Decision-making in the ICU: perspectives of the substitute decision-maker. Intensive Care Med 2003 Jan;29(1):75-82. [doi: 10.1007/s00134-002-1569-y] [Medline: 12528026]
30. Mainetti L, Patrono A, Secco A, Sergi I. An IoT-aware AAL system for elderly people. 2016 Presented at: International Multidisciplinary Conference on Computer and Energy Science (SpliTech); 2016; Split, Croatia p. 1-6. [doi: 10.1109/splitech.2016.7555929]
31. Almeida A, Mulero R, Rametta P, Urošević V, Andrić M, Patrono L. A critical analysis of an IoT—aware AAL system for elderly monitoring. Future Gener Comput Syst 2019 Aug;97:598-619. [doi: 10.1016/j.future.2019.03.019]
32. Pineda D, Urban MC. Inside the Black Blocks. Mowat Centre.: Mowat Centre; 2018. URL: https://munkschool.utoronto.ca/mowatcentre/inside-the-black-blocks/ [accessed 2020-05-20]
33. Zheng X, Mukkamala RR, Vatrapu R, Ordieres-Mere J. Blockchain-based personal health data sharing system using cloud storage. In: IEEE.: IEEE; 2018 Presented at: 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services, Healthcom; 2018; Czech Republic. [doi: 10.1109/healthcom.2018.8531125]
34. Guegan D. Public Blockchain versus Private blockhain. HAL archives. France; 2017. URL: https://halshs.archives-ouvertes.fr/halshs-01524440/document [accessed 2020-05-31]
35. Wright CS. Bitcoin: a peer-to-peer electronic cash system. SSRN J 2008:9 [FREE Full text] [doi: 10.2139/ssrn.3440802]
36. Agbo C, Mahmoud Q, Eklund J. Blockchain technology in healthcare: a systematic review. Healthcare (Basel) 2019 Apr 04;7(2):56 [FREE Full text] [doi: 10.3390/healthcare7020056] [Medline: 30987333]
37. Zheng Z, Xie S, Dai H, Chen X, Wang H. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. 2017 Presented at: Proceedings of the 2017 IEEE 6th International Congress on Big Data; 2017; Honolulu p. 557-564. [doi: 10.1109/bigdatacongress.2017.85]
38. Androulaki E, Barger A, Bortnikov V, Cachin C, Christidis K, de Caro A, et al. Hyperledger fabric: a distributed operating system for permissioned blockchains. In: Proceedings of the Thirteenth EuroSys Conference. 2018 Presented at: EuroSys '18; 2018; Portugal p. 1-15. [doi: 10.1145/3190508.3190538]
39. Bahga A, Madisetti VK. Blockchain platform for industrial internet of things. J Softw Eng Appl 2016;09(10):533-546. [doi: 10.4236/jsea.2016.910036]
40. Underwood S. Blockchain beyond bitcoin. Commun ACM 2016 Oct 28;59(11):15-17. [doi: 10.1145/2994581]
41. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. Bitcoin.: Bitcoin; 2008. URL: https://bitcoin.org/bitcoin.pdf [accessed 2020-05-31]
42. Wang H, Wang Y, Cao Z, Li Z, Xiong G. An overview of blockchain security analysis. Cyber Security 2019:55-72. [doi: 10.1007/978-981-13-6621-5_5]
43. Gorenflo C, Golab L, Keshav S. Mitigating trust issues in electric vehicle charging using a blockchain. In: Proceedings of the Tenth ACM International Conference on Future Energy Systems. 2019 Presented at: e-Energy '19; 2019; USA p. 160-164. [doi: 10.1145/3307772.3328283]
44. IEC - SyC AAL: Active Assisted Living. International Electrotechnical Comission. URL: https://www.iec.ch/dyn/www/f?p=103:214:0::::FSP_ORG_ID,FSP_LANG_ID:11827,25 [accessed 2020-05-31]
45. Morgan-Linnell SK, Stewart DJ, Kurzrock R. US Food and Drug Administration inspections of clinical investigators: overview of results from 1977 to 2009. Clin Cancer Res 2014 Jul 01;20(13):3364-3370 [FREE Full text] [doi: 10.1158/1078-0432.CCR-13-3206] [Medline: 24737548]

46. Statement by Marisol Touraine-Press conference-Clinical trial accident-progress report. Ministry of Solidarity and Health. URL: https://solidarites-sante.gouv.fr/archives/archives-presse/archives-discours/article/intervention-de-marisol-touraine-conference-de-presse-accident-d-essai-clinique [accessed 2020-05-31]
47. Blockchain and GDPR: How blockchain could address five areas associated with GDPR compliance. IBM. 2018. URL: https://iapp.org/resources/article/blockchain-and-gdpr/ [accessed 2020-05-31]
48. Hyperledger Fabric. Hyperledger. URL: https://www.hyperledger.org/use/fabric [accessed 2020-05-23]
49. Introduction. Hyperledger. URL: https://hyperledger-fabric.readthedocs.io/en/release-1.2/whatis.html [accessed 2020-05-23]
50. Hyperledger Fabric Network. Hyperledger. URL: https://hyperledger-fabric.readthedocs.io/en/release-1.2/network/network.html#defining-a-consortium [accessed 2020-05-20]
51. Glossary. Hyperledger. URL: https://hyperledger-fabric.readthedocs.io/en/latest/glossary.html?highlight=organization#organization [accessed 2020-05-31]
52. Peers. Hyperledger. URL: https://hyperledger-fabric.readthedocs.io/en/latest/peers/peers.html [accessed 2020-05-31]
53. The Ordering Service. Hyperledger. URL: https://hyperledger-fabric.readthedocs.io/en/latest/orderer/ordering_service.html?highlight=orderer [accessed 2020-05-22]
54. Membership Service Provider (MSP). Hyperledger. URL: https://hyperledger-fabric.readthedocs.io/en/latest/membership/membership.html [accessed 2020-05-20]
55. Ledger. Hyperledger. URL: https://hyperledger-fabric.readthedocs.io/en/latest/ledger/ledger.html?highlight=world [accessed 2020-05-20]
56. Chaincode for Developers. Hyperledger. URL: https://hyperledger-fabric.readthedocs.io/en/release-1.4/chaincode4ade.html [accessed 2020-05-20]
57. Ethereum Whitepaper. Ethereum. URL: https://ethereum.org/whitepaper/ [accessed 2020-05-20]
58. Bocek T, Rodrigues BB, Strasser T, Stiller B. Blockchains everywhere - A use-case of blockchains in the pharma supply-chain. In: Proceedings of the IM 2017 - 2017 IFIP. 2017 Presented at: IEEE International Symposium on Integrated Network and Service Management; 2017; Portugal URL: http://dl.ifip.org/db/conf/im/im2017exp/119.pdf [doi: 10.23919/inm.2017.7987376]
59. IBM Blockchain Platform. IBM. URL: https://www.ibm.com/blockchain/platform [accessed 2020-05-20]
60. Greenspan G. MultiChain Private Blockchain-White Paper. MultiChain. 2015. URL: https://www.multichain.com/download/MultiChain-White-Paper.pdf [accessed 2020-05-20]
61. HydraChain. GitHub. URL: https://github.com/HydraChain/hydrachain/blob/develop/hc_consensus_explained.md [accessed 2020-05-20]
62. Mcconaghy T, Marques R, Muller A, De Jonghe D, McConaghy T, McMullen G, et al. BigchainDB: A Scalable Blockchain Database - White Paper. MyCourses. 2016. URL: https://mycourses.aalto.fi/pluginfile.php/378362/mod_resource/content/1/bigchaindb-whitepaper.pdf [accessed 2020-05-30]
63. Blockchain network. Hyperledger. URL: https://hyperledger-fabric.readthedocs.io/en/latest/network/network.html [accessed 2020-05-20]
64. Morita PP. Chapter 5 - Design of mobile health technology. In: Sethumadhavan A, Sasangohar F, editors. Design for Health. United States: Academic Press; 2020:87-102.
65. Goyal S, Morita P, Lewis GF, Yu C, Seto E, Cafazzo JA. The systematic design of a behavioural mobile health application for the self-management of type 2 diabetes. Can J Diabetes 2016 Feb;40(1):95-104. [doi: 10.1016/j.jcjd.2015.06.007]
66. Morita PP, Yeung MS, Ferrone M, Taite AK, Madeley C, Stevens Lavigne A, et al. A patient-centered mobile health system that supports asthma self-management (breathe): design, development, and utilization. JMIR Mhealth Uhealth 2019 Jan 28;7(1):e10956 [FREE Full text] [doi: 10.2196/10956] [Medline: 30688654]
67. Sultan M, Kuluski K, McIsaac WJ, Cafazzo JA, Seto E. Turning challenges into design principles: telemonitoring systems for patients with multiple chronic conditions. Health Informatics J 2019 Dec;25(4):1188-1200 [FREE Full text] [doi: 10.1177/1460458217749882] [Medline: 29320911]
68. Omran Y, Henke M, Heines R, Hofmann E. Blockchain-driven supply chain finance: towards a conceptual framework from a buyer perspective. In: International Purchasing and Supply Education and Research Association 2017. 2017 Presented at: IPSERA 2017; 2017; Budapest p. 15-15.
69. Carney D, Çetintemel U, Cherniack M, Convey C, Lee S, Seidman G, et al. Monitoring streams: a new class of data management applications. In: Proceedings of the 28th international conference on Very Large Data Bases. 2002 Presented at: VLDB '02; 2002; China.
70. Fingberg J, Hansen M, Hansen M, Krasemann H, Iacono LL, Probst T, et al. Integrating data custodians in ehealth grids - A digest of security and privacy aspects Digest of Security and Privacy Aspects. In: Informatik 2006. 2006 Presented at: Informatik 2006; 2006; Germany URL: https://dl.gi.de/handle/20.500.12116/23643;jsessionid=3A72C9E5E5E1AE8B8538EB526381958D [doi: 10.1201/9780203486061.ch23]
71. Xia Q, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M. MeDShare: trust-less medical data sharing among cloud service providers via blockchain. IEEE Access 2017;5:14757-14767. [doi: 10.1109/access.2017.2730843]
72. Ongaro D, Ousterhout J. In search of an understandable consensus algorithm. In: Proceedings of USENIX ATC ’14. 2014 Presented at: 2014 USENIX Annual Technical Conference; 2014; USA.

73. Fabric SDK for node.js. Hyperledger. 2018. URL: https://fabric-sdk-node.github.io/index.html [accessed 2020-05-21]
74. Substitute Decision Makers and Naming a Power of Attorney for Personal Care. University Health Network. 2018. URL: https://www.uhn.ca/PatientsFamilies/Health_Information/Health_Topics/Documents/Substitute_Decision_Maker_and_Naming_an_Attorney_for_Personal_Care.pdf [accessed 2020-05-20]
75. Takahashi D. Hu-manity.co uses IBM blockchain to give you the right to control your personal data. VentureBeat. URL: https://venturebeat.com/2018/09/06/hu-manity-co-uses-ibm-blockchain-to-give-you-the-right-to-control-your-personal-data/ [accessed 2020-05-31]
76. Alexandre A. New Bitfury Joint Project to Manage Medical Data Permissions With Blockchain Tech. Cointelegraph. URL: https://cointelegraph.com/news/new-bitfury-joint-project-to-manage-medical-data-permissions-with-blockchain-tech [accessed 2020-05-31]
77. Bitfury announces blockchain-based consent management system; partners with Hancom to distribute Crystal platform. Tokenpost. URL: https://tokenpost.com/Bitfury-announces-blockchain-based-consent-management-system-partners-with-Hancom-to-distribute-Crystal-platform-1603 [accessed 2020-05-31]
78. Zanghi E, Do Coutto Filho MB, Stacchini de Souza JC. Conceptual framework for blockchain-based metering systems. Multiagent Grit Syst 2019 Mar 25;15(1):77-97. [doi: 10.3233/mgs-190303]
79. Building blockchains for a better planet. PwC. 2018. URL: https://www.pwc.com/gx/en/services/sustainability/building-blockchains-for-the-earth.html [accessed 2020-05-20]
80. Siwicki B. The next big thing in pharmacy supply chain: Blockchain. Healthcare IT News. URL: https://www.healthcareitnews.com/news/next-big-thing-pharmacy-supply-chain-blockchain [accessed 2018-10-09]

Abbreviations

AAL: active assisted living
C1: contract one
C2: contract two
DLT: distributed ledger technology
EHR: electronic health record
FDA: Food and Drug Administration
GDPR: General Data Protection Regulation
HF: hyperledger fabric
HFN: hyperledger fabric network
HIPAA: Health Insurance Portability and Accountability Act
IoT: internet of things
MSP: membership service provider
PIPEDA: Personal Information Protection and Electronic Documents Act
RESTful: representation state transfer
RPM: remote patient monitoring
SDK: software development kit
SDM: substitute decision maker

Edited by G Eysenbach; submitted 31.05.20; peer-reviewed by D Istrat, L Luo; comments to author 26.06.20; revised version received 14.08.20; accepted 30.10.20; published 04.12.20.

Please cite as:
Velmovitsky PE, Miranda PADSES, Vaillancourt H, Donovska T, Teague J, Morita PP
A Blockchain-Based Consent Platform for Active Assisted Living: Modeling Study and Conceptual Framework
J Med Internet Res 2020;22(12):e20832
URL: https://www.jmir.org/2020/12/e20832
doi:10.2196/20832
PMID:33275111

Corrigenda and Addenda

Correction: Identifying and Ranking Common COVID-19 Symptoms From Tweets in Arabic: Content Analysis

Eisa Alanazi[1], BSc, MSc, PhD; Abdulaziz Alashaikh[2], BEng, MSc, PhD; Sarah Alqurashi[1], BSc; Aued Alanazi[1], BSc

[1]Center of Innovation and Development in Artificial Intelligence, Umm Al-Qura University, Makkah, Saudi Arabia
[2]University of Jeddah, Jeddah, Saudi Arabia

Corresponding Author:
Eisa Alanazi, BSc, MSc, PhD
Center of Innovation and Development in Artificial Intelligence
Umm Al-Qura University
Prince Sultan Road
Makkah, 11343
Saudi Arabia
Phone: 966 55 601 4171
Email: eaanazi@uqu.edu.sa

Related Article:

Correction of: http://www.jmir.org/2020/11/e21329/

(J Med Internet Res 2020;22(12):e26446) doi:10.2196/26446

In “Identifying and Ranking Common COVID-19 Symptoms From Tweets in Arabic: Content Analysis” (J Med Internet Res 2020;22(11):e21329) one error was noted.

This article was originally published with a layout error causing two inline figures representing Arabic keywords to appear as regular figures in the first paragraph of the Methods section. The sentence where these inline figures should have appeared read as follows:

> *First, we searched Twitter for personal reports of COVID-19 from March 1, 2020, to May 27, 2020, using 2 Arabic keywords Inline Figure 6 and Inline Figure 6, which translate roughly to “I have been diagnosed.”*

This has been corrected to place the two inline figures within this sentence instead of the words “Inline Figure 6” in each instance.

The correction will appear in the online version of the paper on the JMIR Publications website on December 14, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 11.12.20; this is a non–peer-reviewed article;accepted 11.12.20; published 14.12.20.

Please cite as:
Alanazi E, Alashaikh A, Alqurashi S, Alanazi A
Correction: Identifying and Ranking Common COVID-19 Symptoms From Tweets in Arabic: Content Analysis
J Med Internet Res 2020;22(12):e26446
URL: http://www.jmir.org/2020/12/e26446/
doi:10.2196/26446
PMID:33315577

Corrigenda and Addenda

Correction: Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study

Akira-Sebastian Poncette[1,2], MD; Daniel Leon Glauert[1]; Lina Mosch[1]; Katarina Braune[3], MD; Felix Balzer[1,2], MSc, PhD, MD; David Alexander Back[4,5], MD

[1]Department of Anesthesiology and Intensive Care Medicine, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[2]Einstein Center Digital Future, Berlin, Germany

[3]Department of Paediatric Endocrinology and Diabetes, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[4]Department of Traumatology and Orthopaedics, Septic and Reconstructive Surgery, Bundeswehr Hospital Berlin, Berlin, Germany

[5]Dieter Scheffner Center for Medical Education and Educational Research, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

Corresponding Author:
David Alexander Back, MD
Dieter Scheffner Center for Medical Education and Educational Research
Charité – Universitätsmedizin Berlin
Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health
Chariteplatz 1
Berlin, 10117
Germany
Phone: 49 30 2841 1240
Email: david.back@charite.de

Related Article:

Correction of: https://www.jmir.org/2020/10/e22161/

(J Med Internet Res 2020;22(12):e25738) doi:10.2196/25738

In “Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study” (J Med Internet Res 2020;22(10):e22161) the authors noted three errors.

One author's name was displayed as:

> *David Back*

It has now been corrected to:

> *David Alexander Back*

One affiliation for David Alexander Back was listed as:

> *Bundeswehr Hospital Berlin, Clinic for Traumatology and Orthopedics, Septic-Reconstructive Surgery, Berlin, Germany*

It has now been corrected to:

> *Department of Traumatology and Orthopaedics, Septic and Reconstructive Surgery, Bundeswehr Hospital Berlin, Berlin, Germany*

The phone number for corresponding author was incorrectly listed. It has been changed to:

> *49 30 2841 1240*

The correction will appear in the online version of the paper on the JMIR Publications website on December 7, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.11.20; this is a non–peer-reviewed article;accepted 26.11.20; published 07.12.20.

Please cite as:
Poncette AS, Glauert DL, Mosch L, Braune K, Balzer F, Back DA
Correction: Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study
J Med Internet Res 2020;22(12):e25738
URL: https://www.jmir.org/2020/12/e25738
doi:10.2196/25738
PMID:33284785

Corrigenda and Addenda

Correction: Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions

Kimberly D Acquaviva[1], PhD, MSW; Josh Mugele[2], MD; Natasha Abadilla[3], BA; Tyler Adamson[4], MPH; Samantha L Bernstein[5], MSN, RNC-OB, IBCLC; Rakhee K Bhayani[6], MD; Annina Elisabeth Büchi[7], MSc, MD; Darcy Burbage[8], DNP, RN, AOCN, CBCN; Christopher L Carroll[9], MD; Samantha P Davis[10], MS, RRT, RRT-NPS, CHSE; Natasha Dhawan[11], MD; Alice Eaton[12], BSc; Kim English[13], RN, BScN, MN; Jennifer T Grier[14], PhD; Mary K Gurney[15], RPh, PhD, BCPA; Emily S Hahn[16], MD; Heather Haq[17], MHS, MD; Brendan Huang[18], BS; Shikha Jain[19], MD, FACP; Jin Jun[20], RN, PhD; Wesley T Kerr[21], MD, PhD; Timothy Keyes[3], BA; Amelia R Kirby[1], BA; Marion Leary[22], RN, MSN, MPH; Mollie Marr[23], BFA; Ajay Major[24], MD, MBA; Jason V Meisel[25], PMHNP-BC; Erika A Petersen[26], MD, FAANS, FACS; Barak Raguan[27], MD; Allison Rhodes[28], BA; Deborah D Rupert[29,30], MS, MA; Nadia A Sam-Agudu[31,32], MD, CTropMed; Naledi Saul[33], MPM; Jarna R Shah[26], MD; Lisa Kennedy Sheldon[34], PhD, ANP-BC, AOCNP, FAAN; Christian T Sinclair[35], MD; Kerry Spencer[36], PhD; Natalie H Strand[37], MD; Carl G Streed Jr[38], MD, MPH, FACP; Avery M Trudell[39], BSA

[1]School of Nursing, University of Virginia, Charlottesville, VA, United States
[2]Northeast Georgia Medical Center, Gainesville, GA, United States
[3]School of Medicine, Stanford University, Stanford, CA, United States
[4]Center for Public Health and Human Rights, Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD, United States
[5]College of Nursing, Medical University of South Carolina, Charleston, SC, United States
[6]School of Medicine, Washington University, St. Louis, MO, United States
[7]Department of General Internal Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
[8]Oncology Nursing Consultant, Newark, DE, United States
[9]Department of Pediatrics, University of Connecticut, Farmington, CT, United States
[10]Department of Respiratory Care, Boise State University, Boise, ID, United States
[11]Hematology/Oncology Section, Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States
[12]Swansea University Medical School, Swansea, United Kingdom
[13]Trent/Fleming School of Nursing, Peterborough, ON, Canada
[14]Department of Biomedical Sciences, University of South Carolina School of Medicine Greenville, Greenville, SC, United States
[15]College of Pharmacy, Glendale Campus, Midwestern University, Glendale, AZ, United States
[16]Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, United States
[17]Department of Pediatrics, Baylor College of Medicine, Houston, TX, United States
[18]Department of Neurology and Neurosurgery, Tulane University, New Orleans, LA, United States
[19]Division of Hematology and Oncology, University of Illinois Chicago, Chicago, IL, United States
[20]College of Nursing, Ohio State University, Columbus, OH, United States
[21]Department of Neurology, University of California Los Angeles, Los Angeles, CA, United States
[22]School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
[23]School of Medicine, Oregon Health & Science University, Portland, OR, United States
[24]Section of Hematology and Oncology, University of Chicago, Chicago, IL, United States
[25]Hunter School of Nursing, City University of New York, New York, NY, United States
[26]University of Arkansas for Medical Sciences, Little Rock, AR, United States
[27]Meir Medical Center, Kfar Saba, Israel
[28]School of Medicine, Tufts University, Boston, MA, United States
[29]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, United States
[30]State of New York-Stony Brook University, Stony Brook, NY, United States
[31]Institute of Human Virology and Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, United States
[32]International Research Center of Excellence, Institute of Human Virology Nigeria, Abuja, Nigeria
[33]Office of Career and Professional Development, University of California San Francisco, San Francisco, CA, United States

[34]Oncology Nursing Society, Pittsburgh, PA, United States
[35]University of Kansas Health System, Kansas City, KS, United States
[36]Department of Mathematics and Physics, Stevenson University, Owings Mills, MD, United States
[37]Department of Anesthesiology, Mayo Clinic, Phoenix, AZ, United States
[38]Department of Medicine, School of Medicine, Boston University, Boston, MA, United States
[39]McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, United States

Corresponding Author:
Kimberly D Acquaviva, PhD, MSW
School of Nursing
University of Virginia
4005 McLeod Hall
Charlottesville, VA, 22903
United States
Phone: 1 202 423 0984
Email: kda8xj@virginia.edu

Related Article:

Correction of: https://www.jmir.org/2020/12/e25070

(J Med Internet Res 2020;22(12):e26225) doi:10.2196/26225

In “Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions” (J Med Internet Res 2020;22(12):e25070), one of the authors was not mentioned in the original paper.

The following author has been added:

> *Alice Eaton, Bsc*
>
> *Swansea University Medical School, Swansea, United Kingdom*

One of the contributors was missed in the Acknowledgement section. Previously, the Acknowledgement section read as follows:

> *All authors met the ICMJE authorship criteria, including reviewing and approving this manuscript for publication. Each of the 40 coauthors completed an Authorship Attestation Form confirming that they met all 4 of the ICMJE authorship criteria. In addition, each coauthor described on a shared Google Sheet how they made “substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data.” The authors wish to thank Jocelyn C Anderson, PhD, RN, S Alexander Kemery, PhD, RN, and Chrystal L Lewis, PhD, RN for their comments on the draft of this manuscript.*

The corrected Acknowledgement section now reads as follows:

> *All authors met the ICMJE authorship criteria, including reviewing and approving this manuscript for publication. Each of the 40 coauthors completed an Authorship Attestation Form confirming that they met all 4 of the ICMJE authorship criteria. In addition, each coauthor described on a shared Google Sheet how they made “substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data.” The authors wish to thank Jocelyn C Anderson, PhD, RN, S Alexander Kemery, PhD, RN, Rebecca Koszalinski PhD, RN, CRRN, CMSRN, and Chrystal L Lewis, PhD, RN for their comments on the draft of this manuscript.*

The affiliation for authors Brendan Huang and Ajay Major was incorrect. They were displayed as:

> *Brendan Huang*
>
> *Department of Clinical Neurosciences, Tulane University, New Orleans, LA, United States*
>
> *Ajay Major*
>
> *University of Chicago, Section of Hematology and Oncology, Chicago, IL, United States*

They have been updated to:

> *Brendan Huang*
>
> *Department of Neurology and Neurosurgery, Tulane University, New Orleans, LA, United States*
>
> *Ajay Major*
>
> *Section of Hematology and Oncology, University of Chicago, Chicago, IL, United States*

The correction will appear in the online version of the paper on the JMIR Publications website on December 9, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 04.12.20; this is a non–peer-reviewed article;accepted 08.12.20; published 14.12.20.

Please cite as:
Acquaviva KD, Mugele J, Abadilla N, Adamson T, Bernstein SL, Bhayani RK, Büchi AE, Burbage D, Carroll CL, Davis SP, Dhawan N, Eaton A, English K, Grier JT, Gurney MK, Hahn ES, Haq H, Huang B, Jain S, Jun J, Kerr WT, Keyes T, Kirby AR, Leary M, Marr M, Major A, Meisel JV, Petersen EA, Raguan B, Rhodes A, Rupert DD, Sam-Agudu NA, Saul N, Shah JR, Sheldon LK, Sinclair CT, Spencer K, Strand NH, Streed Jr CG, Trudell AM
Correction: Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions
J Med Internet Res 2020;22(12):e26225
URL: http://www.jmir.org/2020/12/e26225/
doi:10.2196/26225
PMID:33315581

©Kimberly D Acquaviva, Josh Mugele, Natasha Abadilla, Tyler Adamson, Samantha L Bernstein, Rakhee K Bhayani, Annina Elisabeth Büchi, Darcy Burbage, Christopher L Carroll, Samantha P Davis, Natasha Dhawan, Alice Eaton, Kim English, Jennifer T Grier, Mary K Gurney, Emily S Hahn, Heather Haq, Brendan Huang, Shikha Jain, Jin Jun, Wesley T Kerr, Timothy Keyes, Amelia R Kirby, Marion Leary, Mollie Marr, Ajay Major, Jason V Meisel, Erika A Petersen, Barak Raguan, Allison Rhodes, Deborah D Rupert, Nadia A Sam-Agudu, Naledi Saul, Jarna R Shah, Lisa Kennedy Sheldon, Christian T Sinclair, Kerry Spencer, Natalie H Strand, Carl G Streed Jr, Avery M Trudell. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.12.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Digital Health Coaching Programs Among Older Employees in Transition to Retirement: Systematic Literature Review

Vera Stara[1*], PhD; Sara Santini[2*], MSc; Johannes Kropf[3], MSc; Barbara D'Amen[2*], PhD

[1]Model of Care and New Technologies, IRCCS INRCA-National Institute of Health and Science on Aging, Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani, Ancona, Italy

[2]Centre for Socio-Economic Research on Aging, IRCCS INRCA-National Institute of Health and Science on Aging, Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani, Ancona, Italy

[3]Health and Environment, Austrian Institute of Technology, Vienna, Austria

[*]these authors contributed equally

Corresponding Author:
Sara Santini, MSc
Centre for Socio-Economic Research on Aging
IRCCS INRCA-National Institute of Health and Science on Aging
Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani
via S. Margherita 5
Ancona, 60124
Italy
Phone: 39 3286191298
Email: s.santini2@inrca.it

Related Article:

Correction of: https://www.jmir.org/2020/9/e17809/

(J Med Internet Res 2020;22(12):e25065) doi:10.2196/25065

In “Digital Health Coaching Programs Among Older Employees in Transition to Retirement: Systematic Literature Review” (J Med Internet Res 2020;22(9):e17809) the authors noted errors in affiliations of three authors.

The affiliation of Vera Stara was originally listed as:

> *Model of Care and New Technologies, National Institute of Health and Science on Aging, Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani, Ancona, Italy*

This has been corrected to:

> *Model of Care and New Technologies, IRCCS INRCA-National Institute of Health and Science on Aging, Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani, Ancona, Italy*

The affiliation of Sara Santini and Barbara D'Amen was originally listed as:

> *Centre for Socio-Economic Research on Aging, National Institute of Health and Science on Aging, Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani, Ancona, Italy*

This has been corrected to:

> *Centre for Socio-Economic Research on Aging, IRCCS INRCA-National Institute of Health and Science on Aging, Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale Ricovero e Cura per Anziani, Ancona, Italy*

This change has also been made under the Corresponding Author address for Sara Santini.

The correction will appear in the online version of the paper on the JMIR Publications website on December 11, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 16.10.20; this is a non–peer-reviewed article;accepted 16.10.20; published 14.12.20.

Please cite as:
Stara V, Santini S, Kropf J, D'Amen B
Correction: Digital Health Coaching Programs Among Older Employees in Transition to Retirement: Systematic Literature Review
J Med Internet Res 2020;22(12):e25065
URL: http://www.jmir.org/2020/12/e25065/
doi:10.2196/25065
PMID:33315575

Viewpoint

Real-Time Communication: Creating a Path to COVID-19 Public Health Activism in Adolescents Using Social Media

Kunmi Sobowale[1], MD; Heather Hilliard[2,3], MHA, MBA; Martha J Ignaszewski[4], MD; Linda Chokroverty[5], MD

[1]Department of Psychiatry and Biobehavioral Sciences, University of California Los Angeles, Los Angeles, CA, United States
[2]Emergency Management, Kenner, LA, United States
[3]School of Professional Advancement, Tulane University, New Orleans, LA, United States
[4]Department of Psychiatry, British Columbia Children’s Hospital, Vancouver General Hospital, Vancouver, BC, Canada
[5]Department of Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, Bronx, NY, United States

Corresponding Author:
Kunmi Sobowale, MD
Department of Psychiatry and Biobehavioral Sciences
University of California Los Angeles
760 Westwood Plaza
Suite 37-384
Los Angeles, CA, 90024
United States
Phone: 1 310 794 7035
Email: osobowale@mednet.ucla.edu

Abstract

The COVID-19 pandemic and related public health efforts limiting in-person social interactions present unique challenges to adolescents. Social media, which is widely used by adolescents, presents an opportunity to counteract these challenges and promote adolescent health and public health activism. However, public health organizations and officials underuse social media to communicate with adolescents. Using well-established risk communication strategies and insights from adolescent development and human-computer interaction literature, we identify current efforts and gaps, and propose recommendations to advance the use of social media risk communication for adolescents during the COVID-19 pandemic and future disasters.

(J Med Internet Res 2020;22(12):e21886) doi:10.2196/21886

KEYWORDS

social media; digital health; COVID-19; adolescent; public health; disaster; communication; affordances

Introduction

The COVID-19 pandemic has affected nearly everyone’s way of life, irrespective of their age. The pandemic has particularly challenged adolescents, who at ages 10-19 years are undergoing some of the most dynamic changes of human development [1]. Individuation from family and connection with peers bring fundamental psychological and social benefits to this age group. The loss of in-person social interactions created by stay-at-home mandates, school closures, and social distancing during COVID-19 has strained these developmental processes. Further, because of the protracted nature of the pandemic and its restrictions, pervasive misinformation on the internet, and incorrect perceptions that healthy adolescents are at low risk for infection [2], some young people have dismissed the dangers of infection and their role in transmission in the greater community to spend time with their peers. We are currently witnessing the unsettling crisis of college campuses reopening this fall in the United States, only to face precipitous shutdowns and quarantines in response to outbreaks of COVID-19, brought about by fraternizing young people [3].

Opportunities exist, however, to help counteract some of these challenges during this uncertain time using adolescent-friendly methods. Given the near ubiquitous use of social media (SM) in this age demographic [4], online platforms traditionally used for entertainment purposes can be mobilized to promote health as well. This viewpoint highlights SM’s potential to serve the critical function of protecting and enabling adolescents during this unprecedented time by providing them with accurate and timely health information that will promote individual engagement in healthy behaviors while mitigating risk. Paralleling community adaptation efforts, SM may promote a paradigm shift from passivity and isolationism toward empowerment and activism.

We focus on well-established risk communication strategies and integrate insights from adolescent development and human-computer interaction literatures to identify current efforts, gaps, and recommendations highlighting two main points. First, to adequately support the promotion of healthy behaviors during COVID-19, public health actors should harness the power of SM to address adolescents through their preferred avenues. Second, content must address adolescent-specific needs to activate adolescents as drivers of health protective behaviors and promoters of community adherence.

Adolescent Use of Social Media

A total of 92% of adolescents use the internet daily and 85% have at least one SM account [4]. Since the start of the COVID-19 pandemic, the use of SM has increased across platforms. For many platforms, particularly those with many adolescent users, the number of new downloads, accounts created, and daily use is at an all-time high [5,6]. In the United States, popular platforms among this age group include YouTube (85%); Instagram (70%); Snapchat (65%) [4]; and, increasingly, TikTok [7]. Because of the ubiquity and stability of adolescent SM use, SM is a promising medium to reach adolescents.

The appeal and frequency of SM use by adolescents can be understood through a developmental perspective. Cognitively, adolescence is marked by a greater ability for self-directed learning and shifts from egocentric thinking to understanding the perspectives of others [8]. In addition to major cognitive changes, adolescence is a time for identity formation and exploration. Socially, adolescents spend more time with peers [9], seeking autonomy from adults. Developmental neuroscience and behavioral science have shown a heightened sensitivity to peer feedback [10,11] and social influences [12] in adolescence. Thus, peer norms and social prestige become more important [1].

The affordances framework provides insight on how SM aligns with these unique aspects of adolescent development [13]. Affordances are key design aspects of SM platforms that underlie how users use them. For adolescents, functional, social, cognitive, emotional, and identity affordances encourage exploration, learning, and practicing of skills that are key for development. Functional affordances refer to features that determine how messages are transmitted and saved for scalability and dissemination, for example, the ability to repost another user’s content. Social affordances are platform features that promote interpersonal interaction. One social affordance is belonging to a group based on a characteristic or shared interest. Identity affordances relate to features allowing users to create and develop their identity and portray it to others. For example, SM profiles can range from platforms that encourage sharing one’s true representation to total anonymity. Cognitive affordances are features that help users learn and share information, for instance, through the use of multimedia such as video as a creative way to promote learning. Emotional affordances are platform features that affect users’ emotional reactions. Examples include the use of personal stories, names, and pictures to create empathy.

Through these affordances, adolescents can explore topics of interest and their identity, connect with others, receive feedback to establish peer norms, and build prestige in a social ecosystem separate from adults. This understanding of adolescent development and the advantages of SM affordance features can inform interventions including risk communication during disasters.

Risk Communication During Disasters and During COVID-19

Studies have identified risk communication, defined as “the exchange of real time information, advice and opinions between experts and people facing threats to their health, economic or social well-being” [14], as an effective public health intervention to inform the public about resources and promote healthy behaviors during disasters. Increasingly, risk communication occurs through SM platforms due to extensive community reach [15]. SM allows the public, and adolescents in particular, to have timely access to a wide variety of topics with dynamic interaction beyond traditional interface media. At the same time, the inaccuracy of disseminated information may undermine an effective response, which has been unfortunately demonstrated in the context of COVID-19 [16].

Established authorities such as local governments and the Centers for Disease Control and Prevention (CDC) provide messages during public health events by using communication patterns that allow SM users to differentiate true facts from misinformation. The CDC has developed a six-principle framework for the dissemination of information during a crisis called Crisis and Emergency Risk Communications (CERC; Textbox 1) [17]. The CDC developed CERC to enable public health response officials to promptly and effectively communicate with the community in emergency situations. CERC is built around psychological and communication sciences, studies in the field of issues management, and learned experiences [17] from previous emergency responses to allow for communicators to work alongside the media and responders to impact community response and recovery from these potentially devastating emergencies. A full response in a crisis can take time, requiring multiple levels of information to be acquired that is incongruent with the need for the community to respond right away. This delay can impair an adequate community response. This is addressed through the basic principles of CERC, which emphasize the importance of being first, being right, and being credible in disaster response.

Textbox 1. Crisis and Emergency Risk Communications principles [17].

Be first

- Communicating information quickly is crucial. Historically, the first source of information often becomes the preferred source.

Be right

- Accuracy establishes credibility. Information can include what is known, what is not known, and actions being implemented to fill in the gaps.

Be credible

- Honesty and truthfulness should not be compromised during crises.

Express empathy

- Suffering should be acknowledged. Addressing emotional reactions and the challenges faced by individuals leads to trust and rapport.

Promote action

- Providing examples of meaningful activities for individuals to engage in may promote calm, reduce anxiety, help restore order, and foster a sense of control [18].

Show respect

- Respectful communication promotes cooperation and rapport.

To further promote information sharing, officials from government, health care organizations, and emergency management systems must reach a variety of listeners through their preferred media channels. The implementation of SM for CERC allows for an additional element of interaction to be incorporated into this framework. Whereas traditional news sources release static, precrafted messages in standard formats, such as press releases, the dynamic nature of SM has allowed for shifts in information dissemination. Historically, management agencies or organizers act as senders transmitting information to the receiving community. Through the use of SM, recipients such as adolescents also take on the role of senders by sharing, commenting, or reposting messages to their networks as a functional affordance.

To facilitate the reading and then the sharing of credible health information, content must quickly create resonance with consumers. Given the intentional brevity of SM posts, resonance must be successfully created with concise messages using few characters and through the provision of meaningful messages containing actionable guidance. Because of the rapid turnover of content, online resonance must create “stickiness,” an aspect of functional affordance, so that individuals access and remain on specific content for a greater duration of time. Stickiness can be extended by further engaging audiences with a familiarity of the rules of SM propagation—for example, the use of hashtags to attract readers. Market research and analytics play crucial roles to ensure that content creation and style meet the expectations of the SM user and lead to selective dispersal of accurate information.

In the CERC framework, directives to the community must evolve through the life cycle of a disaster, defined by four phases: preparedness, response, recovery, and mitigation. During these distinct phases the primary needs are concrete and focused on community safety. Nevertheless, some variables remain consistently vital, such as monitoring for hazards and threats, and flexible communication. Messaging about COVID-19 initially focused on reducing the spread of infection by informing, influencing, and empowering readers to wash their hands, stay at home, social distance, and travel only for essential activities such as picking up groceries or prescriptions. As new information has emerged, the current messaging encourages the use of facial masks in public and ongoing promotion of social distancing as communities begin to reduce restrictions and enable reopening.

Exploring CERC as Applied to Adolescents

Given the pre-existing widespread use of online media and online news sources by the adolescent demographic, CERC messaging through SM may be a particularly effective technique to reach this age group. To effectively engage with adolescents online, messaging must, first, be accurate to be considered an effective public health response. Second, content must be appropriate and relevant to adolescents to be identified as valuable. Here, consideration of the various affordances of SM is important. For adolescents to receive intended communication, it must be posted on platforms that adolescents are routinely accessing. Although the popularity of individual SM platforms will change over time, these affordances will not change and, therefore, can more consistently inform SM interventions. Subsequently, communication on SM that is inclusive, empowering, and promoting appropriate health behaviors can be leveraged as a mechanism to support action by youth and youth activists.

Although national or international health organizations frequently excel in the second and third CERC principles of providing “right” and “credible” information [19], these are time-consuming steps that may prevent adherence to step one of “being first.” By prioritizing speed of information dissemination on SM platforms, organizations and providers may more effectively help adolescents find accurate and up-to-date information from reliable health organizations. Crisis- and disaster-related content directed toward adolescents needs to reflect distilled knowledge from health care providers or public health organizations generated through a funnel process. More simply, this encompasses gathering data, filtering for reliable and consistent information, and then developing content written specifically to promote a selected action. Public health authorities have done little to tailor content on SM platforms for adolescents in prior infectious disease outbreaks [19]. Creating verbiage for adolescent-targeted messaging on SM platforms from health organizations also represents a solution to countering misinformation that has become especially prevalent in the pandemic. In addition, leveraging the functional affordance through ease of replicability of messages and the identity affordance (how adolescents portray themselves online), public health authorities can create precrafted messages for adolescents to repost as their own. This aligns with adolescents' desire for independence from adults and for prestige through social approval. Further, considering the functional affordance of scalability and the importance of peer norms, collaboration with influencers, especially those of similar age, to spread messages will likely be beneficial.

Available research shows that online misinformation on SM platforms, including during natural disasters, spreads faster and reaches more people than true information [20]. This is particularly troubling with the potential for disinformation to lead to significant harm when accurate information from leading health professionals is delayed or absent [21]. Adolescents may be particularly vulnerable to misinformation, using SM as a primary news source [22]. A recent study demonstrated that 25% of the most popular YouTube videos on COVID-19 contain misinformation [21], and 40%-60% of Twitter posts about COVID-19 are driven by bots (autonomous or semiautonomous software applications that mimic humans) [23]. Health care organizations and other public health agencies are tasked with combating this misinformation through CERC principles. In line with the cognitive affordances of knowledge acquisition and self-efficacy, use of interactive and visual media such as videos or quizzes testing knowledge about COVID-19 with corrective information may be more appealing to adolescents than messages with text or static images.

The next three principles of CERC communication ("express empathy," "promote action," "show respect") are equally as important; these require more specific adaptation for the target demographic and directly address the content of messaging. Language must be selected so as not to downplay information or incite anxiety. Word choice must be within the developmental and cognitive ability of the target demographic. Content must reflect the priorities of adolescents, show empathy for their concerns, and be engaging and entertaining [24].

Highlighting the loss of interactions with friends and community, and missing milestone achievements during the pandemic communicates that adolescent voices are heard in the uncertainty of the pandemic. Adolescents should be encouraged to maintain their connectedness with friends and peers by virtual means in absence of in-person contact when possible (encouragement of physical distancing over social distancing). With support of educational agencies, SM platforms have been used creatively toward managing the somber and disappointing cancellations of important rites of passage and celebrations of achievements such as graduations, proms, and summer job opportunities. For many, these letdowns have been softened and brightened through virtual graduations using livestreaming, video, and photo commemorations of the events in festive attire; distribution of digital diplomas; and inspirational speeches by celebrities and dignitaries [25].

These virtual efforts represent attempts to instill normalcy, highlight shared values, and create a sense of security, comfort, and community. In addition, they exemplify the social affordances of belonging and support, and the emotional affordances of using personal stories and ceremonies to generate emotion and garner empathy [13]. Adolescent struggles should not be minimized but, rather, respectfully validated. These efforts promote an alliance between health providers and organizations and adolescents, and make adolescents feel understood. If they feel camaraderie with the organizations and authorities, adolescents will be more likely to identify the importance of the messages and listen to them.

In regard to promoting action, adolescents have traditionally been seen as passive participants during disasters rather than primary agents and collaborators. Adolescents can play an important role in disaster response and preparedness [26]. Peek and colleagues [27] identified many youth contributions in response to Hurricane Katrina. They found that children and adolescents helped raise and donate money, collect material goods and supplies, assist with rebuilding, support behavioral health, develop programs and raise awareness, and found new organizations. Many youth initiatives were embedded within larger organizational responses rather than independently started by children and adolescents, though the authors acknowledge that adolescent-initiated and encouraged activities are important. A few papers highlight youth's use of SM to organize volunteering efforts in response to disasters [28,29].

To our knowledge, there have not been any other similar narrative, systematic, gray, or nonscientific literature reviews for adolescent use of SM in response to disasters including infectious disease outbreaks such as COVID-19. Though research in this area is still nascent, SM use aligning with the affordances framework has been identified as a promising medium to promote prevention in adolescents [30], including of sexually transmitted infections [31,32]. A recent systematic review on the use of technology to scale up youth-led participatory action research found few, albeit positive, examples of the powerful nature of adolescent SM engagement [33]. This suggests that adolescents may possess expertise in SM that has been underrecognized and underused, and there is limited documentation of adolescent contributions on SM platforms [27].

This highlights the importance for health care providers to invest in credible communication directed to adolescents and to support their development as emerging and powerful advocates for improving public health behaviors around the world—particularly at this crucial time in public health history. In line with the United Nations' Sendai Framework for Disaster Risk Reduction 2015-2030 [34] call to include adolescents as agents of change, public health directives should be relayed directly to adolescents.

The best messengers for adolescents are often other adolescents. Some organizations already engage adolescents to participate in response, advocacy, and teaching around important topics in public health. Examples include activation of teen community response teams during disasters (Teen Community Emergency Response Teams [35]) and opportunities for teen leaders to pursue projects that help their communities with preparedness and response during emergencies (the Federal Emergency Management Agency Youth Preparedness Council [36]). Other programs focus on education, empowerment, and building awareness of certain issues (health, youth employment, inequities), accomplished through programs where young people serve as youth ambassadors and youth advocates to the community, such as with Teens Take Charge [37] and the Youth Power Coalition [38]. Enlisting young leaders to promote public health interests allows for this demographic to play an active role in shaping their own future.

Young leaders have already demonstrated great motivation and drive to help others. Content that contains a motivational component can help spread adaptive behaviors to support the community, as reflected through SM posts showing adolescents helping vulnerable neighbors [29]. Social distancing directives in the context of COVID-19 may feel isolating; shifting the schema to highlight the altruism of physical distancing may help to influence desired behaviors at both an individual and population level. This aligns with the adolescents' desire to portray themselves in a positive light. Further, crafting immediate and accurate messages for both youth engagement as well as activation on SM can promote self-discovery, personal development, and local and global connectedness to perpetuate lasting behavior change [39,40].

Using the right communication model can positively impact behavior and change in adolescents during this pandemic. Leveraging SM is an essential part of the disaster communication toolkit to reach adolescents, who are often missed in targeted communications. Enormous potential exists to harness the existing skills of adolescents and the affordances of SM.

Current Risk Mitigation Efforts on Social Media

Although there is evidence supporting the use of the CERC framework's principles during prior disease outbreaks [41,42], the adoption has been slower for adolescents. This may be explained by the fact that many public health and medical organizations previously did not disseminate information through the SM platforms most widely used by youth, such as Instagram, TikTok, Snapchat, and YouTube, or consider affordances [13,22,30,43]. The COVID-19 pandemic has spurred some of these authorities to adapt. For example, the World Health Organization (WHO) started a TikTok account in the wake of COVID-19 and, as of October 2020, has over 2.4 million followers. The CDC has used multiple platforms including YouTube and added a dedicated Snapchat account. The United Nations Children's Fund (UNICEF) has sought to empower adolescents by asking them to share messages and graphics created by UNICEF and the WHO on SM platforms. More organizations and authorities would benefit from following the example of these pioneer agencies to develop SM accounts that reach a more diversely aged target population, especially where adolescents search and receive health information.

The developers of SM platforms should also be commended for identifying creative methods to promote accurate information. Instagram, Twitter, and TikTok verify authoritative accounts, making it easier for users to identify reputable sources to be featured. Similarly, many platforms have dedicated advertisement space to credible organizations. Many platforms including Facebook, Reddit, Twitter, TikTok, and YouTube now recommend or display credible sources of information associated with certain topics or hashtags and may fact-check content [16,44]. In an effort to correct misinformation, some platforms have displayed videos from credible organizations adjacent to COVID-19 videos with misinformed or unsubstantiated claims; however, this juxtaposition may be confusing to young viewers and lead to the conclusion that videos spreading misinformation are legitimate based on their proximity to videos by subject matter authorities. This is especially concerning because SM content recommendation algorithms can lead to a preferential though unintentional selection of misinformation. SM platforms can also demote, suspend, or even remove accounts spreading misinformation; however, this has been a less popular option due to first amendment rights.

Although organizations have developed accounts on SM to directly address youth, content on SM accounts is often geared toward adults and has not been developed specifically for adolescents or fails to acknowledge the unique challenges they face [22]. Few organizations have addressed the understandable reactions to remote schooling and the loss of sports activities and graduation ceremonies. This is a missed opportunity to promote behavior change through emotional and social affordances. Because empathy and respectful communication build trust, this type of messaging is necessary to engage adolescents. Some examples of successful and empathic ways to engage young people have been demonstrated by the United Nations Educational, Scientific and Cultural Organization (UNESCO), UNICEF, and TikTok. UNESCO's #MyCOVID-19Story [45] and UNICEF's #VoicesofYouths [46] campaigns ask youth to describe the challenges their communities face and to share creative and humorous content. TikTok's livestream event for the graduating class of 2020 featured multiple influencers and allowed students to receive digital diplomas. TikTok's prom week (#TikTokProm) [47] allowed adolescents to showcase their social distancing prom experience, which generated over 2 billion views. Since the reopening of universities in the United States, some colleges have expanded the role of paid student influencers (hired for marketing their schools to prospective students) to instead model safe behaviors on SM and connect students to resources as part of efforts to contain recent outbreaks on campuses [48].

In addition to empathy, showing respect is likely to activate adolescents. In some cases, adolescents have been admonished or demonized for engaging in risky behavior. Rather than shaming or being condescending, health organizations should provide positive and corrective experiences. Through functional affordance, more efforts should be made to partner with influencers, especially those with large audiences in the target demographic [22]. Indeed, a number of celebrities have streamed conversations with health authorities. Some of these conversations allow users to actively engage by asking questions in real time. For example, users could ask questions on Netflix's weekly livestream on Instagram called "Wanna Talk About It?" [49], which features conversations between Netflix actors and actresses from popular adolescent shows and mental health professionals to address the mental health and wellness of young people during COVID-19. This interactive way to promote learning and skills exemplifies cognitive affordance.

Many challenges or campaigns have been started to promote action. In general, campaigns try to influence people to engage in a discussion or carry out a set of activities reflecting social affordance. The Vietnamese "Jealous Coronavirus" (Ghen Co Vy) song and the related youth-created #GhenCoVyChallenge dance have been used to promote protective behaviors. The challenge video amassed over 31 million views, and UNICEF and the British Red Cross have shared it [50]. A number of handwashing challenges were created organically and in collaboration with the WHO. The South African Ndlovu Youth Choir created a musical rendition of the WHO's COVID-19 safety advice shared through SM platforms [51]. Proctor and Gamble hired the top influencer on TikTok, an adolescent 15 years of age with nearly 90 million followers, to perform a dance promoting social distancing labeled the #Distancedance [52]. Further, users were encouraged to create and share their own dance. Snapchat has also leveraged its location-sharing app, Zenly [53], to gamify sheltering-in-place. These interactive and visual risk communication strategies are likely more appealing for adolescents, as they leverage cognitive and social affordances and have generated millions of views and shares.

Conclusion

Adolescents represent a vulnerable group who are frequently overlooked in disaster and emergency communication. Given the ongoing COVID-19 pandemic and related future challenges of subsequent waves of infection and vaccination, it is a crucial time for risk communication. As explored in this paper, adolescents can be effectively reached and provide positive benefit given their high level of engagement and resourcefulness through online SM platforms. Many global agencies such as UNICEF have highlighted the individual and collective power of adolescent voices emerging as powerful advocates for change, emphasizing the difference between working for adolescents and working with adolescents.

SM can and should be leveraged by health care leaders and public health organizations to access tools to facilitate dynamic, accurate, and actionable risk communication to ensure awareness of public health issues and best practice guidance for adolescents. Nevertheless, this requires adaptation by health authorities to consider the developmental affordances of SM to reach adolescents and to tailor content to adolescent-specific concerns. Though emerging data showcases online connectedness of adolescents and health authorities leads to adherence to safe behavior and civil activism and advocacy, more research is needed to determine the most effective and safest practices as well as the outcomes of such endeavors. We believe that meeting adolescents *where they are at* on SM ensures that voices of young people are heard and represents an investment in their health and well-being during the COVID-19 pandemic and in future disasters.

Conflicts of Interest

None declared.

References

1. Dahl RE, Allen NB, Wilbrecht L, Suleiman AB. Importance of investing in adolescence from a developmental science perspective. Nature 2018 Feb 21;554(7693):441-450. [doi: 10.1038/nature25770] [Medline: 29469094]
2. Leeb RT, Price S, Sliwa S, Kimball A, Szucs L, Caruso E, et al. COVID-19 trends among school-aged children - United States, March 1-September 19, 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 02;69(39):1410-1415. [doi: 10.15585/mmwr.mm6939e2] [Medline: 33001869]
3. Tracking covid at U.S. colleges and universities. The New York Times. 2020. URL: https://www.nytimes.com/interactive/2020/us/covid-college-cases-tracker.html [accessed 2020-09-08]
4. Anderson M, Jiang J. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-10-01]
5. How teens are coping and connecting in the time of the coronavirus. Common Sense Media. 2020. URL: https://www.commonsensemedia.org/sites/default/files/uploads/pdfs/2020_surveymonkey-key-findings-toplines-teens-and-coronavirus.pdf [accessed 2020-09-16]
6. Koeze E, Popper N. The virus changed the way we internet. The New York Times. 2020. URL: https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html [accessed 2020-06-08]
7. Murphy E, Regan NM, Lavery M, Kim Y, Kumar H. Taking stock with teens - spring 2020 infographic. Piper Sandler. 2020. URL: http://www.pipersandler.com/2col.aspx?id=5957 [accessed 2020-09-20]
8. Dumontheil I, Apperly IA, Blakemore S. Online usage of theory of mind continues to develop in late adolescence. Dev Sci 2010 Mar;13(2):331-338. [doi: 10.1111/j.1467-7687.2009.00888.x] [Medline: 20136929]

9. Lam CB, McHale SM, Crouter AC. Time with peers from middle childhood to late adolescence: developmental course and adjustment correlates. Child Dev 2014;85(4):1677-1693 [FREE Full text] [doi: 10.1111/cdev.12235] [Medline: 24673293]
10. Somerville LH. Special issue on the teenage brain: sensitivity to social evaluation. Curr Dir Psychol Sci 2013 Apr 01;22(2):121-127 [FREE Full text] [doi: 10.1177/0963721413476512] [Medline: 24761055]
11. Sebastian C, Viding E, Williams KD, Blakemore S. Social brain development and the affective consequences of ostracism in adolescence. Brain Cogn 2010 Feb;72(1):134-145. [doi: 10.1016/j.bandc.2009.06.008] [Medline: 19628323]
12. Chein J, Albert D, O'Brien L, Uckert K, Steinberg L. Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. Dev Sci 2011 Mar;14(2):F1-10. [doi: 10.1111/j.1467-7687.2010.01035.x] [Medline: 21499511]
13. Moreno MA, Uhls YT. Applying an affordances approach and a developmental lens to approach adolescent social media use. Digit Health 2019;5:2055207619826678 [FREE Full text] [doi: 10.1177/2055207619826678] [Medline: 30746154]
14. General information on risk communication. World Health Organization. 2015. URL: https://www.who.int/risk-communication/background/en/ [accessed 2020-06-18]
15. Atchison J. How to reach your target audience on social media more effectively. Inc. 2019. URL: https://www.inc.com/young-entrepreneur-council/how-to-reach-your-target-audience-on-social-media-more-effectively.html [accessed 2020-05-15]
16. Limaye RJ, Sauer M, Ali J, Bernstein J, Wahl B, Barnhill A, et al. Building trust while influencing online COVID-19 content in the social media world. Lancet Digit Health 2020 Jun;2(6):e277-e278 [FREE Full text] [doi: 10.1016/S2589-7500(20)30084-4] [Medline: 32322814]
17. CERC: introduction. Centers for Disease Control and Prevention. 2018. URL: https://emergency.cdc.gov/cerc/ppt/CERC_Introduction.pdf [accessed 2020-05-15]
18. Benight CC, Bandura A. Social cognitive theory of posttraumatic recovery: the role of perceived self-efficacy. Behav Res Ther 2004 Oct;42(10):1129-1148. [doi: 10.1016/j.brat.2003.08.008] [Medline: 15350854]
19. Tang L, Bie B, Park S, Zhi D. Social media and outbreaks of emerging infectious diseases: a systematic review of literature. Am J Infect Control 2018 Sep;46(9):962-972 [FREE Full text] [doi: 10.1016/j.ajic.2018.02.010] [Medline: 29628293]
20. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
21. Li HO, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health 2020 May;5(5):e002604 [FREE Full text] [doi: 10.1136/bmjgh-2020-002604] [Medline: 32409327]
22. Robb MB. Teens and the news: the influencers, celebrities, and platforms they say matter most. Common Sense Media. 2020. URL: https://www.commonsensemedia.org/sites/default/files/uploads/research/2020_teensandnews-fullreport_final-release-web.pdf [accessed 2020-09-16]
23. Young VA. Nearly half of the Twitter accounts discussing 'Reopening America' may be bots. Carnegie Mellon School of Computer Science. 2020. URL: https://www.scs.cmu.edu/news/nearly-half-twitter-accounts-discussing-%E2%80%98reopening-america%E2%80%99-may-be-bots [accessed 2020-05-20]
24. Plaisime M, Robertson-James C, Mejia L, Núñez A, Wolf J, Reels S. Social media and teens: a needs assessment exploring the potential role of social media in promoting health. Soc Media Soc 2020 Feb 26;6(1):205630511988602. [doi: 10.1177/2056305119886025]
25. Shafer E. From Oprah to Will Ferrell: all of the 2020 virtual graduation ceremonies. Variety. 2020. URL: https://variety.com/2020/digital/news/oprah-obama-will-ferrell-2020-class-graduation-ceremony-1234607933/ [accessed 2020-05-25]
26. Fothergill A. Children, youth, and disaster. Oxford Res Encyclopedia Nat Hazard Sci 2017:1-28. [doi: 10.1093/acrefore/9780199389407.013.23]
27. Peek L, Austin J, Bittel E, Domingue S, Villarreal M. Children take charge: helping behaviors and organized action among young people after Hurricane Katrina. In: Haeffele S, Storr VH, editors. Bottom-up Responses to Crisis. London: Palgrave Macmillan, Cham; 2020:87-111.
28. Lockwood S, Weaver CK, Munshi D, Simpson M. The self-organising of youth volunteers during the oil spill in New Zealand. N Z J Marine Freshwater Res 2016 Apr 27;50(1):28-41. [doi: 10.1080/00288330.2015.1063515]
29. Lewis KV. The power of interaction rituals: The Student Volunteer Army and the Christchurch earthquakes. Int Small Business J 2013 Oct 18;31(7):811-831. [doi: 10.1177/0266242613478438]
30. Wong CA, Madanay F, Ozer EM, Harris SK, Moore M, Master SO, et al. Digital health technology to enhance adolescent and young adult clinical preventive services: affordances and challenges. J Adolesc Health 2020 Aug;67(2S):S24-S33 [FREE Full text] [doi: 10.1016/j.jadohealth.2019.10.018] [Medline: 32718511]
31. Forsyth R, Purcell C, Barry S, Simpson S, Hunter R, McDaid L, et al. Peer-led intervention to prevent and reduce STI transmission and improve sexual health in secondary schools (STASH): protocol for a feasibility study. Pilot Feasibility Stud 2018;4:180 [FREE Full text] [doi: 10.1186/s40814-018-0354-9] [Medline: 30519482]
32. Dowshen N, Lee S, Matty Lehman B, Castillo M, Mollen C. IknowUshould2: feasibility of a youth-driven social media campaign to promote STI and HIV testing among adolescents in Philadelphia. AIDS Behav 2015 Jun;19 Suppl 2:106-111 [FREE Full text] [doi: 10.1007/s10461-014-0991-9] [Medline: 25563502]
33. Gibbs L, Kornbluh M, Marinkovic K, Bell S, Ozer EJ. Using technology to scale up youth-led participatory action research: a systematic review. J Adolesc Health 2020 Aug;67(2S):S14-S23 [FREE Full text] [doi: 10.1016/j.jadohealth.2019.10.019] [Medline: 32718510]

34. Sendai framework for disaster risk reduction 2015 - 2030. PreventionWeb. 2015. URL: https://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf [accessed 2020-08-08]
35. Teen CERT. Ready. 2019. URL: https://www.ready.gov/teen-cert [accessed 2020-05-25]
36. Youth Preparedness Council. Ready. 2020. URL: https://www.ready.gov/kids/youth-preparedness-council [accessed 2020-05-25]
37. Teens Take Charge. 2020. URL: https://www.teenstakecharge.com/ [accessed 2020-05-25]
38. Youth Power Coalition. 2020. URL: https://www.youthpowercoalition.org/ [accessed 2020-05-25]
39. Tolentino J. What mutual aid can do during a pandemic. The New Yorker. 2020. URL: https://www.newyorker.com/magazine/2020/05/18/what-mutual-aid-can-do-during-a-pandemic [accessed 2020-05-17]
40. Pfefferbaum B, Pfefferbaum RL, Van Horn RL. Involving children in disaster risk reduction: the importance of participation. Eur J Psychotraumatol 2018;9(sup2):1425577 [FREE Full text] [doi: 10.1080/20008198.2018.1425577] [Medline: 29441155]
41. Ophir Y. The effects of news coverage of epidemics on public support for and compliance with the CDC- an experimental study. J Health Commun 2019;24(5):547-558. [doi: 10.1080/10810730.2019.1632990] [Medline: 31244398]
42. Hagen L, Neely S, Scharf R, Keller TE. Social media use for crisis and emergency risk communications during the Zika health crisis. Digital Government Res Pract 2020 Apr 17;1(2):1-21. [doi: 10.1145/3372021]
43. Ruan S, Raeside R, Singleton A, Redfern J, Partridge SR. Limited engaging and interactive online health information for adolescents: a systematic review of Australian websites. Health Commun 2020 Jan 22:1-10. [doi: 10.1080/10410236.2020.1712522] [Medline: 31964190]
44. Waghre P, Seth R. Analysing digital platforms' responses to COVID-19 information disorder. The Takshashila Institution. 2020. URL: https://takshashila.org.in/wp-content/uploads/2020/06/COVID-19-Digital-Platforms-Response-to-Information-Disorder_17-Jun_Takshashila-Discussion-Document_v1.0.pdf [accessed 2020-05-17]
45. My COVID-19 story #YOUTHOFUNESCO. UNESCO. 2020. URL: https://en.unesco.org/youth/my-covid19-story [accessed 2020-05-17]
46. COVID-19 and youth mental health. Voices of Youth. 2020. URL: https://www.voicesofyouth.org/campaign/covid-19-and-youth-mental-health [accessed 2020-05-17]
47. Our promposal for you; dress up and get down with TikTok Prom. TikTok. 2020. URL: https://newsroom.tiktok.com/en-us/our-promposal-for-you-dress-up-and-get-down-with-tiktok-prom [accessed 2020-05-17]
48. Marcus E. Colleges are hiring their own students as Covid-19 safety influencers. The New York Times. 2020. URL: https://www.nytimes.com/2020/09/26/style/coronavirus-campus-influencers.html [accessed 2020-09-03]
49. About Netflix. 2020. URL: https://media.netflix.com/en/press-releases/netflix-and-instagram-partner-for-wanna-talk-about-it-live-q-as [accessed 2020-05-17]
50. Wang K. How a coronavirus safety-themed dance took the world by storm, according to the TikTok star who created it. Business Insider. 2020. URL: https://www.businessinsider.com/coronavirus-song-dance-wash-hands-quang-dang-2020-3 [accessed 2020-05-15]
51. Wickramanayake J. Meet 10 young people leading the COVID-19 response in their communities. United Nations. 2020. URL: https://www.un.org/africarenewal/web-features/coronavirus/meet-10-young-people-leading-covid-19-response-their-communities [accessed 2020-05-15]
52. Pearl D. Adweek. 2020. URL: https://www.adweek.com/agencies/pg-teamed-up-with-tiktok-star-charli-damelio-to-create-a-dance-challenge-for-social-distancing/ [accessed 2020-05-15]
53. Constine J. Snapchat's Zenly launches shelter-in-place leaderboard. TechCrunch. 2020. URL: https://techcrunch.com/2020/03/24/zenly-stay-home/ [accessed 2020-05-15]

Abbreviations

CDC: Centers for Disease Control and Prevention
CERC: Crisis and Emergency Risk Communications
SM: social media
UNESCO: United Nations Educational, Scientific and Cultural Organization
UNICEF: United Nations Children's Fund
WHO: World Health Organization

Edited by G Eysenbach; submitted 28.06.20; peer-reviewed by A Benis, A Silva; comments to author 28.07.20; revised version received 09.10.20; accepted 25.10.20; published 01.12.20.

Please cite as:
Sobowale K, Hilliard H, Ignaszewski MJ, Chokroverty L
Real-Time Communication: Creating a Path to COVID-19 Public Health Activism in Adolescents Using Social Media
J Med Internet Res 2020;22(12):e21886
URL: https://www.jmir.org/2020/12/e21886
doi:10.2196/21886
PMID:33226956

Review

Use of Telehealth During the COVID-19 Pandemic: Scoping Review

Sathyanarayanan Doraiswamy[1*], MD, DHealth, CPH, DipIBLM; Amit Abraham[1*], MD; Ravinder Mamtani[1*], MSc, MD, FACPM, FACOEM, AbOIM, DipABLM; Sohaila Cheema[1*], MPH, CPH, DipIBLM

[1] Institute for Population Health, Weill Cornell Medicine, Doha, Qatar

[*] all authors contributed equally

Corresponding Author:
Sathyanarayanan Doraiswamy, MD, DHealth, CPH, DipIBLM
Institute for Population Health
Weill Cornell Medicine
Education City, Al Luqta St, Ar-Rayyan
Doha, PO 24144
Qatar
Phone: 974 50422860
Email: sdo4003@qatar-med.cornell.edu

Abstract

Background: With over 37.8 million cases and over 1 million deaths worldwide, the COVID-19 pandemic has created a societal and economic upheaval of unparalleled magnitude. A positive transformation has been brought about by innovative solutions in the health care sector that aim to mitigate the impact of COVID-19 on human health. For instance, the use of telehealth has been on the rise amidst this public health emergency.

Objective: Given the unprecedented scale of the pandemic with no definitive endpoint, we aimed to scope the existing telehealth-related literature during a defined period of the ongoing pandemic (ie, January to June 2020).

Methods: Our scoping review was guided by the Joanna Briggs Institute Reviewer Manual. We systematically searched PubMed and Embase databases with specific eligibility criteria. Data extracted from the shortlisted articles included first author and affiliation, journal title, publication type, terminologies used to describe telehealth and their accompanying definitions, health discipline or medical specialties and subspecialties wherein telehealth had been applied, the purpose of telehealth use, and the authors' overall sentiment on telehealth use. We collated the available information and used descriptive statistics to analyze the synthesized data.

Results: In all, 543 articles published across 331 different journals were included in this scoping review. The Journal of Medical Internet Research and its sister journals featured the highest number of articles (25/543, 4.6%). Nearly all (533/543, 98.2%) articles were in English. The majority of the articles were opinions, commentaries, and perspectives (333/543, 61.3%). Most authors of the articles reviewed were from high-income countries (470/543, 86.6%), especially from the United States of America (237/543, 43.6%). In all, 39 different definitions were used to describe terms equivalent to telehealth. A small percentage (42/543, 7.7%) of the articles focused on the provision of COVID-19–related care. Moreover, 49.7% (270/543) of the articles primarily focused on the provision of multiple components of clinical care, and 23% (125/543) of the articles focused on various specialties and subspecialties of internal medicine. For a vast majority (461/543, 84.9%) of the articles, the authors expressed a celebratory sentiment about the use of telehealth.

Conclusions: This review identified considerable emerging literature on telehealth during the first six months of the COVID-19 pandemic, albeit mostly from high-income countries. There is compelling evidence to suggest that telehealth may have a significant effect on advancing health care in the future. However, the feasibility and application of telehealth in resource-limited settings and low- and middle-income countries must be established to avail its potential and transform health care for the world's population. Given the rapidity with which telehealth is advancing, a global consensus on definitions, boundaries, protocols, monitoring, evaluation, and data privacy is urgently needed.

(J Med Internet Res 2020;22(12):e24087) doi:10.2196/24087

KEYWORDS

COVID-19; telehealth; telemedicine; scoping review

Introduction

The ongoing COVID-19 pandemic is a defining moment in the 21st century for many reasons. It has affected over 18.3 million people worldwide and led to over 695,000 deaths, resulting in a societal and economic upheaval of unparalleled magnitude [1]. These unprecedented times have also highlighted the power of science in identifying creative solutions to address this mammoth global challenge. Numerous Information and Communication technology (ICT) tools and innovative approaches, such as tools for online education and telecommuting, were being developed even before the pandemic; these tools gained popularity as people sought to find creative solutions to mitigate the impact of the pandemic [2]. In the health care sector, telehealth or telemedicine practices expanded tremendously during the pandemic and continue to flourish [3].

Telemedicine, a term coined in the 1970s, meant “healing at a distance” [4]. Over the following 4 decades, several peer-reviewed definitions for the term have emerged. In 2007, the World Health Organization (WHO) introduced a standardized definition for telemedicine: “The delivery of healthcare services, where distance is a critical factor, by all healthcare professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of healthcare providers, all in the interests of advancing the health of individuals and their communities” [5]. ICT is defined as a “diverse set of technological tools and resources used to transmit, store, create, share or exchange information. These technological tools and resources include computers, the Internet (websites, blogs, and emails), live broadcasting technologies (radio, television, and webcasting), recorded broadcasting technologies (podcasting, audio and video players and storage devices), and telephony (fixed or mobile, satellite, visio/video-conferencing, etc)” [6].

Although used interchangeably, telehealth by definition refers to health care services involving all health care professions (including education of health care professionals themselves), whereas telemedicine refers to services delivered by physicians only [7]. The last 2 decades have seen the emergence of newer terms such as ehealth, mobile health (mhealth), and digital health, to accommodate more recent advances in ICT-enabled health care [8]. For consistency, and as population health researchers with an interest in the broader health care domain beyond the only physician-delivered medical care model, in this paper, we prefer to use the terminology “telehealth” to refer to all forms of ICT-enabled health care.

Global interest within the scientific community for using telehealth was on the rise even before the COVID-19 pandemic, as evidenced by the increasing number of studies published on this topic in recent years [9]. However, the use of telehealth to improve patient care and population health has been predominantly concentrated among high-income countries rather than low- and low-middle–income countries [10]. Some medical specialties, such as radiology, dermatology, pathology, and psychiatry, used telehealth more frequently than others [7]. Health care professionals express varying sentiments regarding the use of telehealth for patient care. Some consider telehealth as the new holy grail in health care, whereas others are guarded in their opinion about its applicability in the field. Some worry about the lack of face-to-face connection between patients and health care providers, which they believe is needed to develop a therapeutic bond, and others express concern that clinicians are unable to perform all aspects of physical examination while using telehealth [11]. The risk of widening inequity across various population subgroups with the advent of telehealth is also a prevailing concern [12].

With the advancement of ICT in the health care sector to ensure accountability, ethical medical practice, and patient data privacy, several countries imposed legal restrictions and strict regulations regarding the use of this rapidly expanding technology. These factors, in addition to the lack of insurance coverage, were reported as primary barriers to advancing telehealth in many countries prior to the COVID-19 pandemic [13]. With the onset of the pandemic, clinics and hospitals rapidly restricted access for emergency care in order to reduce the risk of disease transmission. To manage patient load, safeguard the health of patients and health care professionals, and ensure continuity of patient care, capable health care systems expanded their health care delivery by providing telehealth services. Moreover, several countries relaxed their laws and regulations pertaining to the use of telehealth. Additionally, with the evolving landscape of health care delivery during the COVID-19 pandemic, insurance companies have now started reimbursing expenses for patient care delivered via telehealth [13].

We hypothesized that there has likely been an increase in the delivery of telehealth-enabled care since the onset of the COVID-19 pandemic, with concurrent experiential reporting by health care professionals who are using telehealth as a health care delivery modality. The application of telehealth to promote health, evaluate and manage diseases, and rehabilitate individuals has been documented in other public health emergencies [14]. Given the unprecedented scale of what we are witnessing in the current COVID-19 pandemic, and its surrounding uncertainties and ramifications on future health care delivery, we aimed to scope the existing literature on telehealth during a defined period in the ongoing pandemic (ie, January to June 2020).

Methods

Study Design

We performed a scoping review consistent with the guidance provided by the Joanna Briggs Institute Reviewer Manual [15,16]. The scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (see Multimedia Appendix 1) [17]. The protocol was registered on the Open Science Framework (registration DOI: 10.17605/OSF.IO/AXN32) [18] on July 19, 2020.

Eligibility Criteria

The eligibility criteria were established a priori. We considered only publications that had been accepted for publication or had been published in peer-reviewed journals. Preprints were not considered. All publication types were considered, including opinions, viewpoints, original research articles, and reviews, with no geographic, time, or language restrictions. Furthermore, we included all publications examining any aspect of telehealth from direct, synchronous videoconferencing between patients and health care providers, to mhealth monitoring via apps, as well as wearable smart devices. We excluded any article whose primary focus was not telehealth, including articles related to molecular studies, modeling studies, and studies that used technology only for a better understanding of disease dynamics with no immediate and direct benefit for health care workers (including medical students and health care managers) or patients.

Search Strategy

We systematically searched 2 electronic databases (PubMed and Embase) from January 1 to June 10, 2020, using both keywords and controlled vocabulary (such as MeSH terms). The search terms were a combination of 2 concepts: (1) COVID-19 and (2) telehealth or telemedicine. The detailed search strategy is provided in Multimedia Appendix 2. A senior information specialist validated the search strategy. For a comprehensive assessment, we also searched the reference lists of all the included articles to identify other studies that may be relevant to our review.

Article Selection and Data Extraction

Articles identified by our search strategy were imported into Rayyan, the online systematic review software, and duplicates were removed [19]. SD screened the title and abstracts of the identified articles. AA checked the excluded studies and was able to confirm that the exclusion criteria were correctly applied. Subsequently, SD and AA individually extracted data from 50% of the included studies each. We developed a standardized Microsoft Excel (Microsoft Corp.) template for data extraction to tabulate specific information from the included studies, such as journal title, written language of the article, reference to telehealth or its variants in the article title, reference to COVID-19 or its variants in the article title, publication type, country of the first author's affiliation, country of focus of the article, terminologies used to describe telehealth and their accompanying definitions, the purpose of telehealth use, and the health discipline or medical specialty and subspecialties wherein telehealth had been applied during the study period.

Additionally, for each article, we categorized the overall sentiment expressed by the authors about the usefulness of telehealth. This categorization was based on the framework developed by Nettleton et al [20] and subsequently used by Dol et al [21] in their scoping reviews of the use of internet technology and social media in medical, sociological, and popular literature. Accordingly, the sentiments were categorized as "celebratory" (ie, authors provide a positive appraisal of telehealth use during the pandemic), "contingent" (ie, authors recognize the potential positive contribution but also acknowledge its potential limitations), or "concerned" (ie, authors identify challenges and caution on the imbalance that telehealth may create in health care delivery).

Finally, SD and AA randomly cross-checked approximately 10% of each other's extractions to ensure correctness and completeness of the extracted data. No discrepancies were noted.

Data Analysis and Synthesis

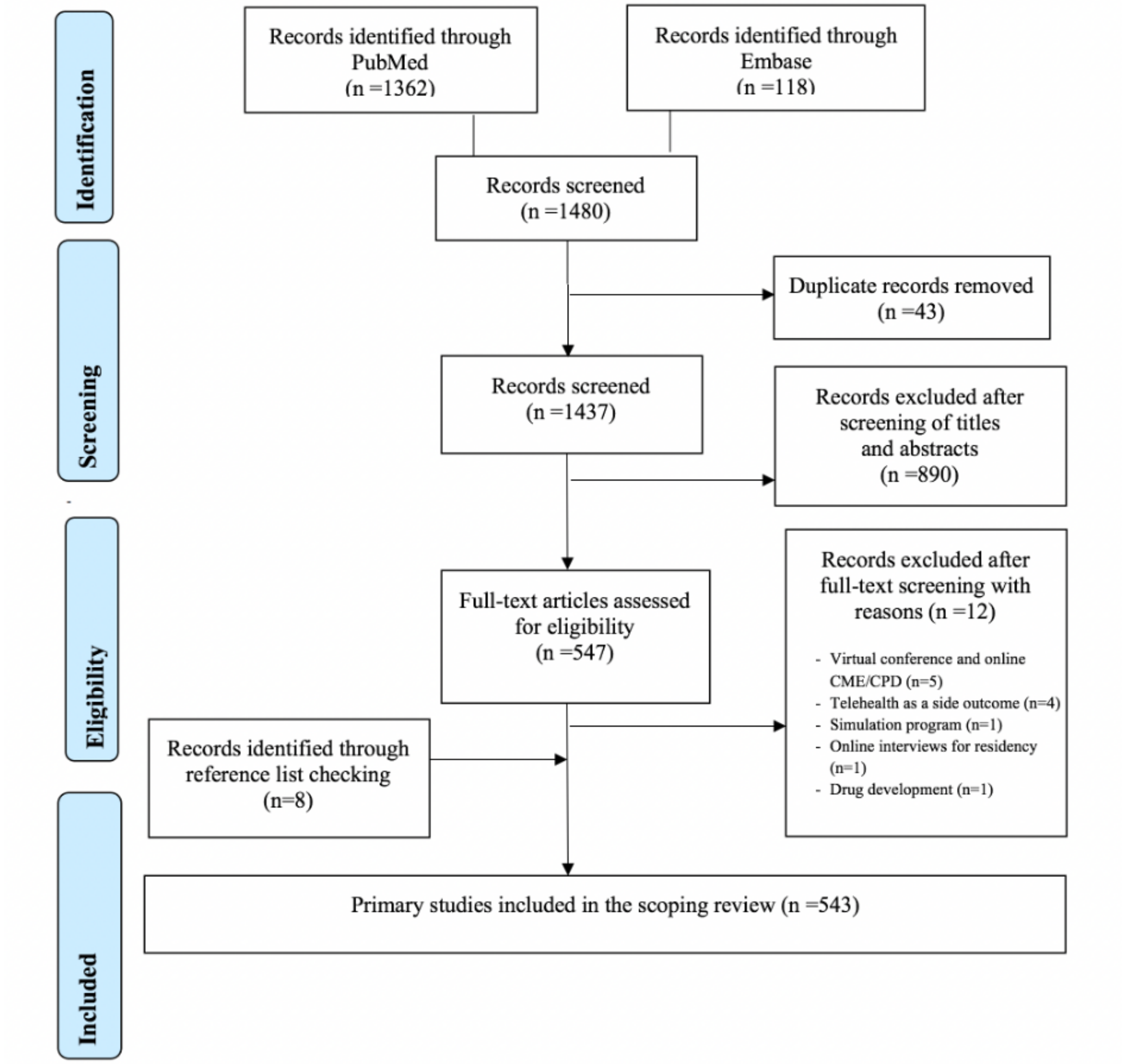
We synthesized the collated data by using descriptive statistics (frequencies and proportions). We used Microsoft Excel and SPSS software (version 26.0; IBM Corp) to analyze the data.

Results

Selection of Articles

After removal of duplicates, we identified 1437 articles in our initial search. Next, based on our eligibility criteria, articles were excluded at various screening stages: 890 at the title and abstract screening stage and 12 at the full-text screening stage. We also searched the reference lists of the included articles and found another 8 relevant articles for inclusion. Thus, a total of 543 articles were included in our review. Figure 1 shows the PRISMA flow chart illustrating the publication selection process. The full list of included studies is provided as Multimedia Appendix 3.

Figure 1. PRISMA 2009 flowchart. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses. CME: continuing medical education. CPD: continuing professional development.



Characteristics of Included Articles

The 543 articles included in our review were published across 331 journals. The *Journal of Medical Internet Research* and its sister journals featured the highest number of articles (25/543, 4.6%), followed by *Telemedicine and e-Health* (19/543, 3.5%). Other journals featuring a high number of included articles were the *Journal of Diabetes Science and Technology* (18/543, 3.3%), *Journal of the American Academy of Dermatology* (9/543, 1.7%), *Otolaryngology-Head and Neck Surgery* (9/543, 1.7%), and the *Journal of Rural Health* (8/543, 1.5%). The remaining journals featured fewer than 8 articles on telehealth in the context of COVID-19. Nearly all (533/543, 98.2%) of the included articles were in English. A small number of articles (8/543, 1.5%) were in French and Spanish, and 1 article each was in German and Hungarian. While the authors of this study read English and French, information from the Spanish, German, and Hungarian articles was extracted from their English abstracts by using Google Translate.

Our manual search of only the titles of the articles included in our study identified that 72.4% (393/543) of the articles made a direct reference to telehealth or its variants and 95.6% (519/543) of them made a direct reference to COVID-19 or its variants. A majority (333/543, 61.3%) of the articles were published as an opinion, commentary, and perspective, followed by empirical research (63/543, 11.6%) and review (narratives or systematic review and meta-analyses: 33/543, 6.1%). In all, 5% (27/543) of the articles were categorized as “others” (ie, Bridging the gap, Business horizons, Care delivery, Clinical forum, Clinical practice guidelines, Consensus statement, Curb side consult, First view, How to do it, Leaders focus, Orthopedic forum, Practice, Practice guidelines, Practice & policy, Special feature, and Training room). Article-type categorization was unclear for 16% (87/543) of the articles.

The first authors of the articles included in our review were affiliated (place of work) in 42 different countries. Most first authors were from the United States of America (237/543, 43.6%), followed by much a smaller number from the United Kingdom, Italy, India, Canada, Australia, France, China, Spain, and Singapore, in decreasing order (Table 1). The geographic focus of the articles also varied, with 32.8% (178/543) of all articles focusing on the United States; this was closely followed by 28.5% (155/543) of all articles had a global or regional focus (Table 2). The geographical focus of the articles largely matched the countries of affiliations of the first authors. We further grouped the countries according to the World Bank's income status [22] and WHO's regional classification [23]. The complete lists of countries based on the above classifications are provided in Multimedia Appendix 4. The vast majority of articles were published from high-income countries (470/543, 86.6%; World Bank classification) and the Americas region (277/543, 51%; WHO classification), closely followed by Europe (168/543, 30.9%). A summary of this comparison is provided in Figures 2 and 3.

Table 1. Country of affiliation of the first authors (N=543).

Country of affiliation[a]	Number of articles, n (%)
USA	237 (43.6)
UK	52 (9.6)
Italy	44 (8.1)
India	25 (4.6)
Canada	25 (4.6)
Australia	16 (2.9)
France	15 (2.8)
China	15 (2.8)
Spain	13 (2.4)
Singapore	11 (2.0)

[a] The first author's place of work

Table 2. Geographic focus of published articles included in the review (N=543).

Country of focus[a]	Number of articles, n (%)
USA	178 (32.8)
Global[b]	155 (28.5)
UK	33 (6.1)
Italy	31 (5.7)
India	19 (3.5)
Canada	18 (3.3)
France	12 (2.2)
Australia	11 (2.0)
China	11 (2.0)
Spain	9 (1.7)
Brazil	9 (1.7)
Iran	6 (1.1)
Germany	6 (1.1)

[a] Some articles focused on more than one country.

[b] Articles that had a global focus, covering more than 3 countries.

Figure 2. Number of articles based on the World Bank's classification of countries by income level (N=543).

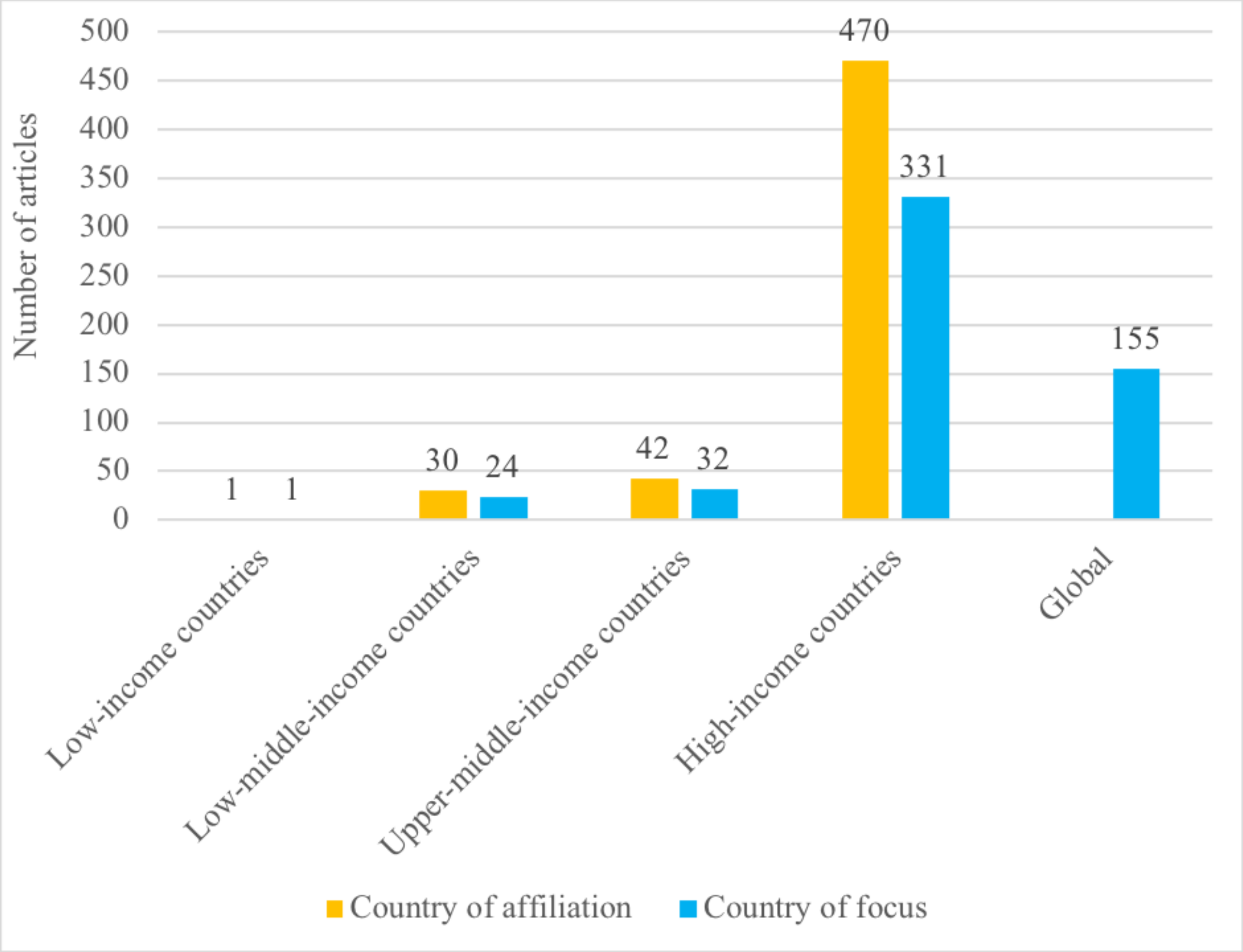
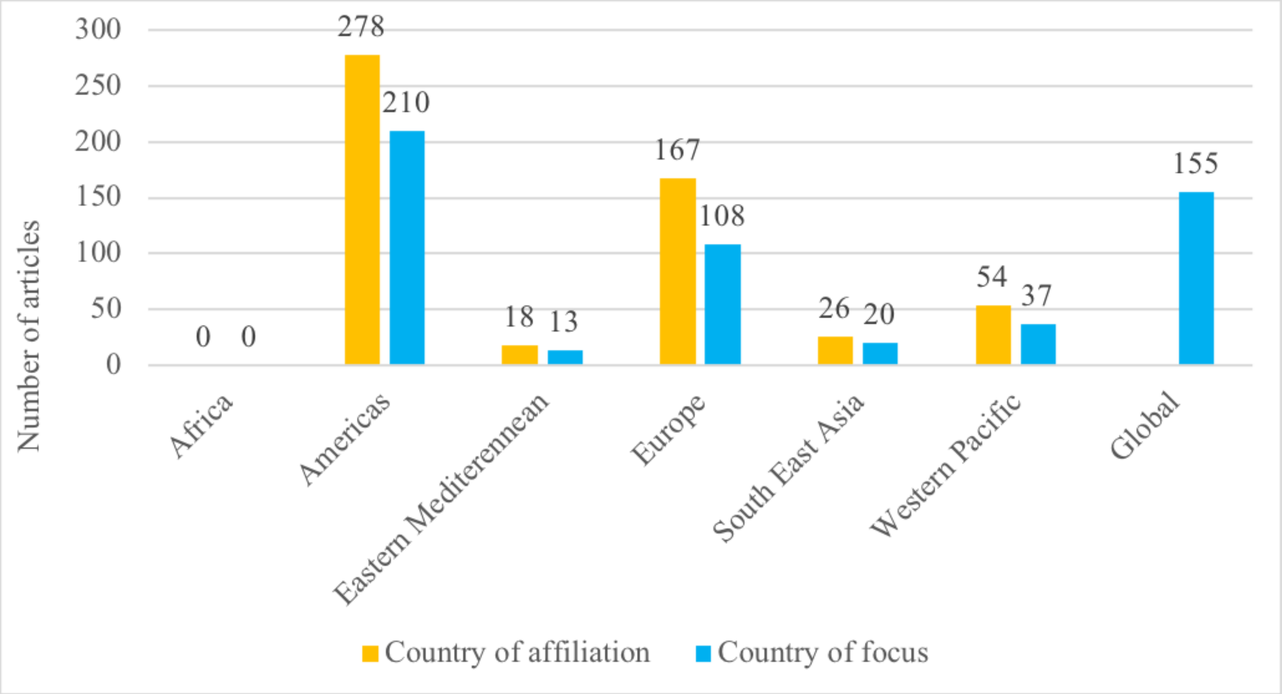


Figure 3. Number of articles based on the World Health Organization's regional classification (N=543).



Terminologies Used and Their Definitions

We found 39 different terminologies related to telehealth in the articles included in our review. The most frequent terminologies used were "telehealth" and "telemedicine." Additionally, frequent use of tele-prefix specialty or subspecialty (eg, teleneurology, telestroke), digital health, ehealth, remote health, video visits, and others was observed. Articles also included references to emerging ICT concepts such as artificial intelligence (AI), robotics, and wearable technology in the context of telehealth. Of the 543 articles, 105 (19.3%) articles included a definition of telehealth-related terminologies. Of these, 13 articles attempted to define telehealth, whereas 52 articles used varying definitions for telemedicine, only 2 of which matched the definition put forth by the WHO. The various definitions of "telehealth" and its variants provided in the included articles have been compiled and presented in Multimedia Appendix 5.

Purpose of Telehealth Use

Of the articles included in our review, 7.7% (42/543) articles focused on the provision of telehealth related to COVID-19, whereas 92.3% (501/543) reported provision of health care support for conditions not related to COVID-19 (eg, tuberculosis, HIV, diabetes, and stroke). We found that the actual purpose of telehealth use during the COVID-19 pandemic varied across studies (Table 3), with the most common purpose being providing multiple components of clinical care (270/543, 49.7%), including any combination of triage, diagnosis, treatment, follow-up, and rehabilitation services. Other purposes for telehealth use were follow-up care (83/543, 15.3%), medical education (54/543, 9.9%), diagnosis only (39/543, 7.2%), and rehabilitation (24/543, 4.4%).

Table 3. Various purposes of telehealth use during the COVID-19 pandemic (N=543).

Purpose	Number of articles, n (%)
Clinical care	270 (49.7)
Follow-up	83 (15.3)
Medical education	54 (9.9)
Diagnosis	39 (7.2)
Rehabilitation	24 (4.4)
Health communication	20 (3.7)
Triage	19 (3.5)
Surveillance or contact tracing	16 (2.9)
Research	12 (2.2)
Health care worker wellbeing	6 (1.1)

Telehealth Use in Various Medical Specialties and Subspecialties

With regard to the specialties covered in the articles included in our review, 89.9% (488/543) of the articles discussed the application of telehealth in medicine and dentistry, and 9.9% (54/543) of the articles focused on medical education. Among the 54 articles on medical education, 10 articles discussed residency training, 8 discussed undergraduate medical student training, and the remaining 36 discussed telehealth in the context of medical education in general (covering undergraduate, postgraduate, residency training, and continuing professional development programs). Only 1 article discussed the application of telehealth in dentistry. We further classified the specialties and subspecialties of medicine covered by the articles using the framework proposed by the American Association of Medical Colleges on various specialties and subspecialties of medicine [24]. We found that 12.9% (70/543) of the articles focused on the use of telehealth for medicine in general with no reference to any specialty. The majority of the remaining articles focused on telehealth use in the following medical specialties: internal medicine (125/543, 23%), preventive medicine (56/543, 10.3%), psychiatry (42/543, 7.7%), surgery (36/543, 6.6%), neurology (33/543, 6.1%), otolaryngology (23/543, 4.2%), and dermatology (23/543, 4.2%). Additional analysis of the subspecialties revealed that the top 5 subspecialties deploying telehealth were endocrinology (30/543, 5.5%), oncology (25/543, 4.6%), geriatrics (23/543, 4.2%), cardiovascular (20/543, 3.7%), and orthopedics (10/543,1.8%). The numbers of articles grouped across various specialties and subspecialties of medicine are reported in Table 4.

Table 4. Number of articles included in the review categorized according to various medical specialties, subspecialties, and specific disease conditions (N=543).

Medical specialty[a], subspecialty[b], and specific disease or condition[c]	Number of articles, n (%)
Internal medicine	125 (23)
Endocrinology	30 (5.5)
Diabetes mellitus/gestational diabetes	9 (1.7)
Diabetes insipidus	1 (0.2)
Hyponatremia	1 (0.2)
Thyroid conditions	1 (0.2)
Eating disorders	1 (0.2)
Oncology	25 (4.6)
Head and neck	8 (1.5)
Lung	3 (0.6)
Neurosurgical oncology	2 (0.4)
Radiation oncology	2 (0.4)
Surgical oncology	1 (0.2)
Myelofibrosis	1 (0.2)
Melanoma	1 (0.2)
Prostate and other urological cancers	1 (0.2)
Cardio-oncology	1 (0.2)
Geriatrics	23 (4.2)
Physical activity	1 (0.2)
Cardiovascular	20 (3.7)
Heart failure	4 (0.7)
Aortic stenosis	1 (0.2)
Vascular	1 (0.2)
Gastroenterology	9 (1.7)
Inflammatory bowel disease	5 (0.9)
Hepatology	4 (0.7)
Rheumatology	9 (1.7)
Systemic lupus erythematosus	2 (0.4)
Systemic sclerosis	1 (0.2)
Rheumatoid arthritis	1 (0.2)
Pulmonology	4 (0.7)
Cystic fibrosis	2 (0.4)
Asthma	1 (0.2)
Severe respiratory failure (ECMO[d])	1 (0.2)
Critical care	3 (0.6)
Hematology	2 (0.4)
Hemophilia	2 (0.4)
No particular specialty mentioned[e]	70 (12.9)
Psychiatry	42 (7.7)
Substance use	4 (0.7)
Counselling	4 (0.7)

Medical specialty[a], subspecialty[b], and specific disease or condition[c]	Number of articles, n (%)
Mood disorders	1 (0.2)
Eating disorders	1 (0.2)
Preventive medicine	56 (10.3)
COVID-19	42 (7.7)
Health education	1 (0.2)
Noncommunicable diseases	1 (0.2)
Surgery	36 (6.6)
Orthopedics	10 (1.8)
Neurosurgery	8 (1.5)
Transplant	3 (0.6)
Orofacial	2 (0.4)
Trauma	1 (0.2)
Thoracic	1 (0.2)
Plastic	1 (0.2)
Neurology	33 (6.1)
Epilepsy	5 (0.9)
Stroke	4 (0.7)
Amyotrophic lateral sclerosis/motor neuron disease	3 (0.6)
Parkinson’s disease and movement disorders	3 (0.6)
Migraine	2 (0.4)
Dementia	2 (0.4)
Multiple sclerosis	1 (0.2)
Demyelinating diseases	1 (0.2)
Otolaryngology	23 (4.2)
Pediatric ENT	2 (0.4)
Dysphagia and swallowing disorders	2 (0.4)
Speech pathology/laryngology	1 (0.2)
Speech apnea	1 (0.2)
Dermatology	22 (4.1)
Psoriasis	1 (0.2)
Dermatosis	1 (0.2)
Chronic inflammatory skin diseases	1 (0.2)
Atopic dermatitis	1 (0.2)
Cutaneous lesions	1 (0.2)
Acne	1 (0.2)
Pediatrics	15 (2.8)
Gastroenterology	3 (0.6)
Neonatology	2 (0.4)
Well-baby clinic	1 (0.2)
Neurology	1 (0.2)
Rehabilitation	1 (0.2)
Inherited metabolic diseases	1 (0.2)

Medical specialty[a], subspecialty[b], and specific disease or condition[c]	Number of articles, n (%)
Cleft palate/lip	1 (0.2)
Overweight/obese children	1 (0.2)
Adolescent health/eating disorders	1 (0.2)
Adolescent health/ADHD	1 (0.2)
Obstetrics and gynecology	13(2.4)
Feto-maternal medicine	6 (1.1)
Antenatal care	2 (0.4)
Female pelvic medicine and reconstructive surgery	2 (0.4)
Physical medicine and rehabilitation	11(2.0)
Physical therapy	4 (0.7)
Musculoskeletal pain	1 (0.2)
Ophthalmology	10 (1.8)
Glaucoma	2 (0.4)
Oculoplastic conditions	1 (0.2)
Urology	10 (1.8)
Infectious disease	7 (1.3)
HIV	6 (1.1)
Tuberculosis	1 (0.2)
Hospice and palliative medicine	6 (1.1)
Diagnostic radiology	4 (0.7)
Ultrasound	1 (0.2)
Anesthesiology	2 (0.4)
Anatomical and clinical pathology	
Digitalization of diagnostic services	1 (0.2)
Allergies and immunology	2 (0.4)
Allergies	2 (0.4)

[a] Number of articles in each medical specialty; where not mentioned, the article is included under the general practice.

[b] Number of articles in each subspecialty; where not mentioned, the article is included in the specialty category only.

[c] Number of articles discussing each specific disease or condition; where not mentioned, the article is included in subspecialty category only.

[d] ECMO: extracorporeal membranous oxygenation.

[e] Articles discussed various aspects of medicine without reference to any specific medical specialty.

In the articles included in our review, the top 5 diseases or conditions for which telehealth was used were diabetes mellitus (9/543, 1.7%), head and neck cancers (8/543, 1.5%), HIV (6/543, 1.1%), epilepsy (5/543, 0.9%), and inflammatory bowel disease (5/543, 0.9%). Within the preventive medicine specialty, 7.7% (42/543) of the articles discussed the various forms of COVID-19 prevention, treatment, and control.

Classification of Sentiments on Telehealth Use

As described in the Methods, we categorized the sentiments expressed by the authors based on the framework proposed by Nettleton et al [20] and subsequently used by Dol et al [21]. The majority (461/543, 84.9%) of articles were celebratory in nature, followed by those that were contingent (74/543, 13.6%), and concerned (8/543, 1.5%). Articles that were categorized as contingent and concerned predominantly stated the need for more evidence on (1) patient satisfaction, (2) cost-effectiveness, (3) efficacy and accuracy of care, and (4) health equity.

Discussion

Principal Findings

The findings from our scoping review indicate that substantial published literature on telehealth has emerged and continues to do during the ongoing COVID-19 pandemic. The eagerness of health care providers and expert researchers to share their opinions and research findings on the application and future potential of telehealth is evident. We observed that telehealth remains a topic of interest for a wide variety of journals (generic and specialized). This is not surprising because telehealth not only shows promise but also has the potential to improve health care access globally [25].

The vast majority of published articles in the literature are in the English language. This is also true for articles featuring technology-related information, as often there are no suitable words in native languages to define technological advancements. This gap in communication can consequently be a deterrent to publishing, with the views of experts and researchers from non–English-speaking countries being discounted or overlooked [26]. It is interesting to note that most of the articles included in our review had a direct reference to COVID-19 and telehealth in their titles, emphasizing the value and use of telehealth during the pandemic. Although journals usually prefer to publish empirical research and reviews, the fact that the vast majority of the articles in our review were viewpoints and opinions demonstrates the willingness of journals to publish such articles while empirical research on the topic continues to emerge during the pandemic. These experiences, arguments, and debates can help identify future research questions. Although the majority of shorter communications such as opinions, commentaries, and viewpoints may be useful for future research, studies have found that these article types are often not backed by adequate data and/or are poorly reviewed in a rush to disseminate relevant scientific knowledge [27]. The evolving nature of the COVID-19 pandemic necessitates swift publishing; however, the measures to protect scientific integrity cannot be overemphasized [28]. Any follow-up systematic reviews on telehealth during the pandemic should place high emphasis on the quality of studies included. Moreover, the small proportion of empirical research and reviews on telehealth during this period should also be seen as a call for additional scientifically robust primary studies with hard data and statistics to offer current and reliable evidence on telehealth.

Our analysis of the first authors’ affiliations and the study’s geographic focus showed that a vast majority of the publications originated from high-income and upper-middle–income countries in the Americas, Europe, and the Western Pacific regions. The higher number of publications from the United States of America is commensurate with the recent rapid surge in telehealth use seen in that country. This can be attributed to the flexibility provided by the Health Insurance Portability and Accountability Act of 1996 and the willingness of insurance companies to reimburse for the services provided via telehealth [29]. Among the low- and upper-middle–income countries, the geographic focus was on India, China, Brazil, and Iran. These countries are large, have been profoundly affected by the COVID-19 pandemic, and had some telehealth-related infrastructure in place even before the onset of the pandemic. We found that the list of countries with the maximum number of publications, based on our review, was mostly consistent with the country-wise publication output as published by the National Science Board of the United States of America [30]. Interestingly, we found no publication originating from Africa; this may be because COVID-19 has had a relatively low overall impact on health within the continent to date [31] and the fact that telehealth infrastructure availability is limited in many African countries [32].

The wide range and variation in the definitions of telehealth used by the authors, despite WHO’s efforts to standardize the definition, is concerning. This is reflective of the lack of consensus in the scientific world on what constitutes telehealth [8,33]. Given more recent developments in the field, including the growing scope of wearable technology and AI, that can augment telehealth and compensate for some of its limitations (eg, physical examination and continuous monitoring) [34], there is a need to revisit the definition of telehealth and arrive at a global consensus. We believe that the use of very broad terminologies (such as ehealth) or very narrow terminologies (such as telestroke) might hamper the standardization and introduction of legal and regulatory provisions to facilitate the use of telehealth. Uniformity in terminology is also important for future evidence-generating systematic reviews, such as those evaluating the efficiency and effectiveness of telehealth.

Telehealth and COVID-19

Our study demonstrates that telehealth has been used broadly in the context of the COVID-19 pandemic as an aid to the active management of patients with COVID-19; for surveillance, triage, and diagnosis; treatment including e-prescriptions; follow-up care; and rehabilitation. It is interesting to note that telehealth has been complemented by the use of wearable devices and selfcare equipment, such as glucometers, handheld blood pressure monitors, pulse oximeters, and digital stethoscopes [35,36]. The use of such equipment, as necessitated by the pandemic, has favorably augmented telehealth use. This can be expected to continue to serve as an adjunct to the provision of telehealth and in-person health care delivery in the long-term [37]. Futuristic advancements in the development and deployment of wearable devices and unobtrusive sensing systems in telehealth offer considerable scope for potential applications and research [35].

Our findings suggest that telehealth has been extensively utilized for medical education during the COVID-19 pandemic. Its use must have considerably helped medical schools in the delivery and continuity of medical education and training. It must also have allowed students to keep on track so that they are able to complete graduation requirements in a timely manner despite challenges associated with the pandemic. Telehealth has been previously used for case rounds and case discussions in residency training [38]. Case discussions, in particular, involve an intersection between tele-education and telehealth, as they bring expert (clinical) educators closer to trainees and students based in remote locations [39]. Certain medical specialties such as dermatology, neurology, obstetrics and gynecology, orthopedics, pathology, psychiatry, and surgery were more likely to use telehealth in residency training than other specialties [40]. Moreover, almost all medical specialties have tapped into telehealth for medical care. Specialties such as dermatology, pathology, and psychiatry have long employed telehealth to provide services to patients [7], whereas other specialties such as surgery, anesthesiology, and oncology are beginning to find telehealth useful at least to deliver certain components of their regular service.

Our findings suggest telehealth has been used to manage a wide spectrum of noncommunicable and communicable diseases, including COVID-19. An earlier scoping review had identified the predominant use of telehealth for noncommunicable diseases, but it did not focus on communicable diseases [41].

In addition to supporting disease-specific management, telehealth has also been extended to provide holistic medical care to specific target groups such as adolescents and older people [42,43].

Given the large scope of telehealth services analyzed by our study, it was no surprise to find that authors of an overwhelming majority of published articles expressed a celebratory sentiment regarding the use of telehealth. This finding is in line with that of many past reviews on the subject [9,44-46]. Authors of the few articles that expressed contingent and concerned sentiments cited the need to review additional evidence on the use of telehealth for health care delivery and to further explore ethics and equity in the context of telehealth use [47-51]. Given that the majority of articles were opinions and perspectives, this is a concern that cannot be overlooked.

Study Limitations

Although this scoping review was conducted according to the suggested methodology, we acknowledge our study has some limitations. We searched only 2 databases and did not actively search the grey literature and preprints. Consequently, our search may not have been exhaustive. Furthermore, following best practices in scoping reviews would have required us to have 2 independent reviewers involved in the screening and data extraction or charting stages. Given the time sensitivity and high volume of publications, an optimized approach had to be considered. However, we have ensured transparency by clearly outlining the process followed in the Methods section. Although the initial title and abstract screening was performed by 1 reviewer, given the large number of studies for data charting, the full-text screening for data extraction was completed by 2 reviewers. Any variation between the 2 reviewers would have been mitigated by the standardized, well-defined, self-explanatory data extraction form used. As most of the data (except for the sentiment analysis) extracted from the articles contained factual information (such as the name of the journal, type of publication, country of affiliation of first author) drawn directly from the articles and were objectively verifiable, the likelihood of variation remained low. The fact that both the reviewers cross-verified at least 10% of each other’s work provides an added level of cushion to the process. Another limitation to this scoping review is the rapidity with which articles seem to be published. From the time of our search to the writing of the manuscript, the number of publications in PubMed alone had almost doubled. As a result, this review is expected to serve the purpose of being an interim scoping review only and can be further updated as the need arises.

Conclusions

Our scoping review highlights the exponential use of telehealth during a defined period of the ongoing COVID-19 pandemic. Experiential reports pertaining to telehealth use are being published extensively, albeit mostly from high-income countries such as the United States, in particular. A wide variety of journals, including specialty journals, are increasingly publishing more material on telehealth. This is in tandem with an increasing number of medical specialties beginning to use telehealth for patient care. Our study also found many subspecialties of medicine that utilize telehealth. Emerging technologies, including wearable devices and AI, are futuristic adjuncts to telehealth, which may help mitigate some of its limitations. The positive sentiment expressed by most authors regarding the use of telehealth is reflective of the developing enthusiasm and receptiveness for this technology. However, we cannot overlook the need for additional robust evidence on the safety and effectiveness of telehealth as compared to the traditional health care delivery model, as pointed out by some authors. Our scoping review demonstrates the breadth and depth of data currently being generated in this area and will enable future systematic reviews and meta-analyses to help address research gaps and answer emerging questions.

Telehealth may have a significant effect in advancing health care in the future. If it has the potential to transform health care, we must ensure that low- and middle-income countries can benefit from it. More North-South and South-South collaborations between academics and practitioners are needed to establish the feasibility and utility of telehealth in resource-limited settings. The advances in medical education facilitated through telehealth are noteworthy. The school curricula for health professionals must be reformed to include specific education pertaining to health care delivery utilizing telehealth, as this can have a profound effect on patient outcomes and the overall health of the population. Continuing medical education and continuing professional development in telehealth use must also concurrently be offered to practicing health care practitioners, so they can become cognizant of and comfortable with using this modality to aid the provision of health care delivery. Given the rapidity with which telehealth is advancing, a global consensus on definitions, boundaries, protocols, monitoring, evaluation, and data privacy are urgently needed.

Acknowledgments

We would like to thank Ms. Sinead M O’Rourke, Content Development Specialist, Distributed eLibrary, Weill Cornell Medicine, Qatar, for her assistance in manuscript editing.

Authors' Contributions

All authors collectively contributed to the conception of the study. SD and AA were involved in the literature search, screening, and data extraction. SD conducted the analyses and wrote the manuscript with support from AA, SC, and RM. All authors read, edited, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.
[PDF File (Adobe PDF File), 515 KB - jmir_v22i12e24087_app1.pdf]

Multimedia Appendix 2

Search strategy utilized in the scoping review.
[DOCX File , 15 KB - jmir_v22i12e24087_app2.docx]

Multimedia Appendix 3

List of published articles included in the scoping review.
[PDF File (Adobe PDF File), 228 KB - jmir_v22i12e24087_app3.pdf]

Multimedia Appendix 4

Classification of the country of affiliation of the first author and the country of focus of the study.
[PDF File (Adobe PDF File), 90 KB - jmir_v22i12e24087_app4.pdf]

Multimedia Appendix 5

Various definitions of telehealth and its variants used in the studies included in the scoping review.
[DOC File , 54 KB - jmir_v22i12e24087_app5.doc]

References

1. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int [accessed 2020-11-23]
2. Baig A, Hall B, Jenkins P, Lamarre E, McCarthy B. The COVID-19 recovery will be digital: a plan for the first 90 days. McKinsey Digital. 2020 May 14. URL: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-covid-19-recovery-will-be-digital-a-plan-for-the-first-90-days# [accessed 2020-11-23]
3. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
4. Strehle EM, Shabde N. One hundred years of telemedicine: does this new technology have a place in paediatrics? Arch Dis Child 2006 Dec;91(12):956-959 [FREE Full text] [doi: 10.1136/adc.2006.099622] [Medline: 17119071]
5. A health telematics policy - in support of WHO's Health-for-all strategy for global health development. Report of the WHO Group Consultation on Health Telematics, 11-16 December, Geneva, 1997. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/63857/WHO_DGO_98.1.pdf?sequence=1&isAllowed=y [accessed 2020-11-23]
6. Guide to measuring information and communication technologies (ICT) in education. UNESCO Institute for Statistics. URL: http://uis.unesco.org/sites/default/files/documents/guide-to-measuring-information-and-communication-technologies-ict-in-education-en_0.pdf [accessed 2020-11-23]
7. Telemedicine: opportunities and developments in member states - Report on the second global survey on eHealth (Global Observatory for eHealth Series, 2). Geneva, Switzerland: World Health Organization; 2010.
8. Scott R, Mars M. Here we go again - 'digital health'. J Int Soc Telemed eHealth 2019 Jan 22;7(e1):1-2. [doi: 10.29086/JISfTeH.7.e1]
9. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/nejmra1601705]
10. Scott R, Mars M. Telehealth in the developing world: current status and future prospects. SHTT 2015 Feb:25. [doi: 10.2147/shtt.s75184]
11. Ellimoottil C, An L, Moyer M, Sossong S, Hollander JE. Challenges and opportunities faced by large health systems implementing telehealth. Health Aff (Millwood) 2018 Dec;37(12):1955-1959. [doi: 10.1377/hlthaff.2018.05099] [Medline: 30633667]
12. Kaplan B, Litewka S. Ethical challenges of telemedicine and telehealth. Camb Q Healthc Ethics 2008 Oct 01;17(4):401-416. [doi: 10.1017/s0963180108080535]
13. Cason J, Brannon JA. Telehealth regulatory and legal considerations: frequently asked questions. Int J Telerehabil 2011;3(2):15-18 [FREE Full text] [doi: 10.5195/ijt.2011.6077] [Medline: 25945185]
14. Keshvardoost S, Bahaadinbeigy K, Fatehi F. Role of telehealth in the management of COVID-19: lessons learned from previous SARS, MERS, and Ebola outbreaks. Telemed J E Health 2020 Jul;26(7):850-852. [doi: 10.1089/tmj.2020.0105] [Medline: 32329659]

15. Peters M, Godfrey C, McInerney P. Chapter 11: Scoping Reviews (2020 version). In: JBI Manual for Evidence Synthesis. Adelaide, Australia: The Joanna Briggs Institute; 2017.
16. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. International Journal of Evidence-Based Healthcare 2015;13(3):141-146. [doi: 10.1097/xeb.0000000000000050]
17. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/m18-0850]
18. Open Science Framework (. OSF home. 2020. URL: https://osf.io/axn32 [accessed 2020-06-02]
19. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
20. Nettleton S, Burrows R, O'Malley L. The mundane realities of the everyday lay use of the internet for health, and their consequences for media convergence. Sociol Health Illn 2005 Nov;27(7):972-992 [FREE Full text] [doi: 10.1111/j.1467-9566.2005.00466.x] [Medline: 16313525]
21. Dol J, Tutelman PR, Chambers CT, Barwick M, Drake EK, Parker JA, et al. Health researchers' use of social media: scoping review. J Med Internet Res 2019 Nov 13;21(11):e13687 [FREE Full text] [doi: 10.2196/13687] [Medline: 31719028]
22. World Bank Country and Lending Groups. The World Bank. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2020-07-15]
23. WHO regional offices. World Health Organization. URL: https://www.who.int/about/who-we-are/regional-offices [accessed 2020-07-15]
24. Specialty Profiles. American Association of Medical Colleges. URL: https://www.aamc.org/cim/explore-options/specialty-profiles [accessed 2020-08-17]
25. Thimbleby H. Technology and the future of healthcare. J Public Health Res 2013 Dec 01;2(3):e28 [FREE Full text] [doi: 10.4081/jphr.2013.e28] [Medline: 25170499]
26. Powell K. Publishing: foreign tongues. Nature 2012 Jul 05;487(7405):129-131. [doi: 10.1038/nj7405-129a] [Medline: 22768394]
27. Horbach SPJM. Pandemic publishing: medical journals strongly speed up their publication process for COVID-19. Quantitative Science Studies 2020 Aug;1(3):1056-1067. [doi: 10.1162/qss_a_00076]
28. Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN, Lazarus JV. Pandemic publishing poses a new COVID-19 challenge. Nat Hum Behav 2020 Jul;4(7):666-669. [doi: 10.1038/s41562-020-0911-0] [Medline: 32576981]
29. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
30. Karen W. Publication Output: US Trends and International Comparisons. National Science Board - Science and Engineering Indicators. 2019. URL: https://ncses.nsf.gov/pubs/nsb20206/ [accessed 2020-06-02]
31. Nachega J, Seydi M, Zumla A. The late arrival of coronavirus disease 2019 (COVID-19) in Africa: mitigating pan-continental spread. Clin Infect Dis 2020 Jul 28;71(15):875-878 [FREE Full text] [doi: 10.1093/cid/ciaa353] [Medline: 32227121]
32. Wamala D, Augustine K. A meta-analysis of telemedicine success in Africa. J Pathol Inform 2013;4:6 [FREE Full text] [doi: 10.4103/2153-3539.112686] [Medline: 23858382]
33. Fatehi F, Wootton R. Telemedicine, telehealth or e-health? A bibliometric analysis of the trends in the use of these terms. J Telemed Telecare 2012 Dec;18(8):460-464. [doi: 10.1258/jtt.2012.gth108] [Medline: 23209265]
34. Kuziemsky C, Maeder AJ, John O, Gogia SB, Basu A, Meher S, et al. Role of artificial intelligence within the telehealth domain. Yearb Med Inform 2019 Aug;28(1):35-40 [FREE Full text] [doi: 10.1055/s-0039-1677897] [Medline: 31022750]
35. Ding X, Clifton D, Ji N, Lovell NH, Bonato P, Chen W, et al. Wearable sensing and telehealth technology with potential applications in the coronavirus pandemic. IEEE Rev Biomed Eng Preprint posted online on May 11, 2020. [doi: 10.1109/RBME.2020.2992838] [Medline: 32396101]
36. Klum M, Urban M, Tigges T, Pielmus AG, Feldheiser A, Schmitt T, et al. Wearable cardiorespiratory monitoring employing a multimodal digital patch stethoscope: estimation of ECG, PEP, LVET and respiration using a 55 mm single-lead ECG and phonocardiogram. Sensors (Basel) 2020 Apr 04;20(7):2033 [FREE Full text] [doi: 10.3390/s20072033] [Medline: 32260436]
37. Greiwe J, Nyenhuis SM. Wearable technology and how this can be implemented into clinical practice. Curr Allergy Asthma Rep 2020 Jun 06;20(8):36 [FREE Full text] [doi: 10.1007/s11882-020-00927-3] [Medline: 32506184]
38. Kanneganti A, Sia C, Ashokka B, Ooi SBS. Continuing medical education during a pandemic: an academic institution's experience. Postgrad Med J 2020 Jul;96(1137):384-386. [doi: 10.1136/postgradmedj-2020-137840] [Medline: 32404498]
39. Gonzales-Zamora JA, Alave J, De Lima-Corvino DF, Fernandez A. Videoconferences of infectious diseases: an educational tool that transcends borders. A useful tool also for the current COVID-19 pandemic. Infez Med 2020;28(2):135-138 [FREE Full text] [Medline: 32275254]
40. Dedeilia A, Sotiropoulos MG, Hanrahan JG, Janga D, Dedeilias P, Sideris M. Medical and surgical education challenges and innovations in the COVID-19 Era: a systematic review. In Vivo 2020 Jun;34(3 Suppl):1603-1611. [doi: 10.21873/invivo.11950] [Medline: 32503818]
41. Abbott PA, Liu Y. A scoping review of telehealth. Yearb Med Inform 2018 Mar 05;22(01):51-58. [doi: 10.1055/s-0038-1638832]

42. Chen S. An online solution focused brief therapy for adolescent anxiety during the novel coronavirus disease (COVID-19) pandemic: a structured summary of a study protocol for a randomised controlled trial. Trials 2020 May 13;21(1):402 [FREE Full text] [doi: 10.1186/s13063-020-04355-6] [Medline: 32404126]
43. Noone C, McSharry J, Smalle M, Burns A, Dwan K, Devane D, et al. Video calls for reducing social isolation and loneliness in older people: a rapid review. Cochrane Database Syst Rev 2020 May 21;5:CD013632. [doi: 10.1002/14651858.CD013632] [Medline: 32441330]
44. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
45. McLean S, Sheikh A, Cresswell K, Nurmatov U, Mukherjee M, Hemmi A, et al. The impact of telehealthcare on the quality and safety of care: a systematic overview. PLoS One 2013;8(8):e71238 [FREE Full text] [doi: 10.1371/journal.pone.0071238] [Medline: 23977001]
46. Wootton R. Twenty years of telemedicine in chronic disease management--an evidence synthesis. J Telemed Telecare 2012 Jun 06;18(4):211-220 [FREE Full text] [doi: 10.1258/jtt.2012.120219] [Medline: 22674020]
47. Romanick-Schmiedl S, Raghu G. Telemedicine - maintaining quality during times of transition. Nat Rev Dis Primers 2020 Jun 01;6(1):45 [FREE Full text] [doi: 10.1038/s41572-020-0185-x] [Medline: 32483168]
48. Rosen C, Joffe S, Kelz R. COVID-19 moves medicine into a virtual space: a paradigm shift from touch to talk to establish trust. Ann Surg 2020 Aug;272(2):e159-e160 [FREE Full text] [doi: 10.1097/SLA.0000000000004098] [Medline: 32675525]
49. Sousa AD, Karia S. Telepsychiatry during COVID-19: some clinical, public health, and ethical dilemmas. Indian J Public Health 2020 Jun;64(Supplement):S245-S246 [FREE Full text] [doi: 10.4103/ijph.IJPH_511_20] [Medline: 32496267]
50. Crowe M, Inder M, Farmar R, Carlyle D. Delivering psychotherapy by video conference in the time of COVID-19: Some considerations. J Psychiatr Ment Health Nurse Epub ahead of print posted online on May 15, 2020. [doi: 10.1111/jpm.12659] [Medline: 32413164]
51. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]

Abbreviations

AI: artificial intelligence
ICT: Information and Communication Technology
mhealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews
WHO: World Health Organization

Edited by G Eysenbach; submitted 03.09.20; peer-reviewed by LS Hanjani, M Sotiropoulos, C Noone; comments to author 09.10.20; revised version received 26.10.20; accepted 30.10.20; published 01.12.20.

Please cite as:
Doraiswamy S, Abraham A, Mamtani R, Cheema S
Use of Telehealth During the COVID-19 Pandemic: Scoping Review
J Med Internet Res 2020;22(12):e24087
URL: https://www.jmir.org/2020/12/e24087
doi:10.2196/24087
PMID:33147166

Original Paper

Development and External Validation of a Machine Learning Tool to Rule Out COVID-19 Among Adults in the Emergency Department Using Routine Blood Tests: A Large, Multicenter, Real-World Study

Timothy B Plante[1,2], MD, MHS; Aaron M Blau[2], MD; Adrian N Berg[3,4], BS; Aaron S Weinberg[5], MD, MPhil; Ik C Jun[5], MD; Victor F Tapson[5], MD; Tanya S Kanigan[4], PhD; Artur B Adib[4], PhD

[1]Larner College of Medicine at the University of Vermont, Colchester, VT, United States
[2]University of Vermont Medical Center, Burlington, VT, United States
[3]Larner College of Medicine at the University of Vermont, Burlington, VT, United States
[4]Biocogniv Inc, South Burlington, VT, United States
[5]Cedars-Sinai Medical Center, Los Angeles, CA, United States

Corresponding Author:
Timothy B Plante, MD, MHS
Larner College of Medicine at the University of Vermont
360 S Park Drive
Suite 206B
Colchester, VT, 05446
United States
Phone: 1 802 656 3688
Email: timothy.plante@uvm.edu

Abstract

Background: Conventional diagnosis of COVID-19 with reverse transcription polymerase chain reaction (RT-PCR) testing (hereafter, PCR) is associated with prolonged time to diagnosis and significant costs to run the test. The SARS-CoV-2 virus might lead to characteristic patterns in the results of widely available, routine blood tests that could be identified with machine learning methodologies. Machine learning modalities integrating findings from these common laboratory test results might accelerate ruling out COVID-19 in emergency department patients.

Objective: We sought to develop (ie, train and internally validate with cross-validation techniques) and externally validate a machine learning model to rule out COVID 19 using only routine blood tests among adults in emergency departments.

Methods: Using clinical data from emergency departments (EDs) from 66 US hospitals before the pandemic (before the end of December 2019) or during the pandemic (March-July 2020), we included patients aged ≥20 years in the study time frame. We excluded those with missing laboratory results. Model training used 2183 PCR-confirmed cases from 43 hospitals during the pandemic; negative controls were 10,000 prepandemic patients from the same hospitals. External validation used 23 hospitals with 1020 PCR-confirmed cases and 171,734 prepandemic negative controls. The main outcome was COVID 19 status predicted using same-day routine laboratory results. Model performance was assessed with area under the receiver operating characteristic (AUROC) curve as well as sensitivity, specificity, and negative predictive value (NPV).

Results: Of 192,779 patients included in the training, external validation, and sensitivity data sets (median age decile 50 [IQR 30-60] years, 40.5% male [78,249/192,779]), AUROC for training and external validation was 0.91 (95% CI 0.90-0.92). Using a risk score cutoff of 1.0 (out of 100) in the external validation data set, the model achieved sensitivity of 95.9% and specificity of 41.7%; with a cutoff of 2.0, sensitivity was 92.6% and specificity was 59.9%. At the cutoff of 2.0, the NPVs at a prevalence of 1%, 10%, and 20% were 99.9%, 98.6%, and 97%, respectively.

Conclusions: A machine learning model developed with multicenter clinical data integrating commonly collected ED laboratory data demonstrated high rule-out accuracy for COVID-19 status, and might inform selective use of PCR-based testing.

(J Med Internet Res 2020;22(12):e24048) doi:10.2196/24048

KEYWORDS

COVID-19; SARS-CoV-2; machine learning; artificial intelligence; electronic medical records; laboratory results; development; validation; testing; model; emergency department

Introduction

SARS-CoV-2 is the cause of COVID-19, which continues to spread in an uncontrolled manner across the United States [1]. COVID-19 management includes patient isolation and supportive care [2]. This strategy requires expeditious COVID-19 diagnosis, but components required for the reverse transcription polymerase chain reaction (RT-PCR; hereafter, PCR) assay have been reported to be in short supply in some locations during the pandemic, leading to delays in results [3]. In the absence of a widely available PCR test with rapid turnaround, there is an urgent need to identify alternative means for stratifying risk of patients seeking care during the COVID-19 pandemic.

Risk assessment models might identify those at low risk of active COVID-19 using available data from the clinical encounter [4,5]. In contrast to traditional model-building techniques, machine learning technologies consider complex linear and nonlinear associations between independent variables and identify characteristic patterns of commonly collected data among patients with COVID-19 [6]. A test with high sensitivity and diagnostic yield (ie, fraction of patients ruled out) could be used in a manner analogous to other rule-out tests, such as D-dimer for pulmonary embolism [7].

Using emergency department (ED) patient encounters from a well-established multicenter clinical database, we sought to describe the development of a machine learning model for ruling out COVID-19 using only routinely collected laboratory tests. Furthermore, we aimed to assess the area under the receiver operating characteristic (AUROC) curve of a machine learning model's concordance with both COVID-19 PCR test results (for positives) and prepandemic patients (for negatives). We hypothesized that such a machine learning model would enable the ruling out of the disease with sensitivity >90% and diagnostic yield >50%.

Methods

Study Design and Setting

This analysis and its reporting is compliant with the Standards for Reporting Diagnostic Accuracy Studies (STARD) statement [8]. This cross-sectional study was performed using 3 data sets of deidentified, patient-level electronic medical records of adult patients in an ED. The Premier Healthcare Database (PHD) is a large database of 1041 US hospitals from all 9 US geographic regions defined by the US Census [9]. At time of writing, 155 hospitals contribute SARS-CoV-2 RNA testing results to the PHD. We separately obtained data from Cedars-Sinai Medical Center (CSMC), an 886-bed academic medical center in Los Angeles, CA, and the Beth Israel Deaconess Medical Center (BIDMC), a 673-bed academic medical center in Boston, MA. An inclusion flow diagram and descriptions of these data sets are provided in Section A of Multimedia Appendix 1.

Prepandemic and Pandemic Time Frames

Two time frames were used, defined by the date of ED visit: prepandemic (before January 2020) and pandemic (March 2020 through July 2020). January 2020 and February 2020 were not included due to the lack of widespread monitoring or diagnostic tests for COVID-19 in the United States during this time frame, even though SARS-CoV-2 community transmission was present in the United States during this time [10]. Clinical encounter data from the PHD were available for the prepandemic (January 2019-December 2019) and pandemic (March 2020 through July 2020) time frames. CSMC data were available for patients with COVID-19 during the pandemic time frame only (March-April 2020). BIDMC data were available across an extended prepandemic time frame (2008-2019) only for patients who were admitted through the ED.

Selection of Participants

Eligible patient encounters (hereafter, patients) were adults aged ≥20 years in an ED at an included center during one of the prepandemic or pandemic time frames. Patients were excluded if they were missing a laboratory result included in the model on the day of presentation to the ED or if any of their laboratory results were reported with inappropriate units or incorrect specimen type. Patients were defined as PCR-positive for COVID-19 (hereafter, PCR-positive) if they had a positive SARS-CoV-2 RNA test on the day of presentation to the ED. We chose PCR rather than antigen positivity to define the cases as PCR is commonly used as the reference standard in COVID-19 diagnosis [11,12].

Training Population and Definition of COVID-19 Cases and Controls

Training occurred in the PHD database only. The PHD training and external validation sets were split by hospital, and only hospitals that reported COVID positives as well as the blood tests required for the model were included in the analysis (64 total). Of these, 43 hospitals were randomly assigned to the training set, and 21 to the external validation set (hereafter, PHD holdout). Cases came from the pandemic time frame, and any patients in this time frame without a positive PCR test were excluded. Contemporary COVID-19 PCR assays have elevated false negative rates, which could lead to mislabeled data and hence to degraded model performance [13]. Due to this, prepandemic controls randomly selected from the 43 PHD hospitals in the training set were used in place of PCR-negative patients during the pandemic.

External Validation Populations

The external validation data set used 3 data sources: 952 PCR-positives and 154,341 prepandemic visits from the 21 hospitals in the PHD holdout set; 68 PCR-positive patients from CSMC; and 17,393 prepandemic (2008-2019) patient encounters from BIDMC. Patients in the pandemic time frame without a positive PCR test were excluded. All prepandemic patients were treated as negatives when evaluating the performance of the model in predicting COVID-19 status. The prepandemic patients from the PHD holdout were chosen so as to match the top 20 most frequent primary diagnoses given to patients without COVID-19 during the pandemic, as coded by Clinical Classifications Software Refined (CCSR) codes (listed in Section B of Multimedia Appendix 1).

Sensitivity Analysis Population

To evaluate how the model generalizes to pandemic time frame patients only, we performed a sensitivity analysis using patients presenting to the ED in the 21 centers from the PHD holdout with any SARS-CoV-2 PCR result available on day of presentation. This differed from the other analyses as negatives were from the same time frame as the positives. This resulted in a total of 952 PCR-positive patients and 6890 PCR-negative patients in the pandemic period (March-July 2020).

Subgroup Analyses

The AUROC was tabulated by decile of age, sex, race, admission or discharge status, and intensive care unit (ICU) admission status in the external validation data set. The distribution of risk scores was also visualized for all studied cohorts through box plots. For PCR-positives, this included positives from CSMC, and PHD visits that had a single positive PCR result as well as visits that had a negative result before a positive result (both on the day of presentation). For PCR negatives during the pandemic, this included patients with both single- and double-PCR results on the day of presentation. For prepandemic encounters, the scores for all eligible BIDMC patients were considered, as well as those from the PHD holdout that matched the top 20 CCSR (non–COVID-19) codes observed during the pandemic.

Model Development (Training and Internal Validation With Cross-Tabulation Techniques)

The model was intended to estimate COVID-19 status on the day of presentation to an ED using common laboratory tests collected that day. Model training began with 29 routinely measured features (ie, potential or included model covariates) comprising the comprehensive metabolic panel and the complete blood count with differential. Recursive feature elimination with cross-validation (RFECV) was performed to arrive at the final 15 features [14]. We used the gradient boosting model as implemented in XGBoost [15] for all results. No hyperparameter optimization was performed and default parameters were used. Performance on the training set was evaluated through stratified 5-fold cross-validation. Performance in the external validation and sensitivity analysis data sets was obtained after training the model on the entire training set.

Statistical Analysis

Baseline demographics, ED disposition, and included laboratory features from the training, external validation, and sensitivity analysis data sets were tabulated by COVID-19 status. Visualization of the distribution of features used box plots, ordered by feature importance (compare with list values in Section C of Multimedia Appendix 1). Model discrimination was visualized with receiver operating characteristic (ROC) curves and estimation of the AUROC. AUROC 95% CIs were estimated with bootstrapping. Hosmer-Lemeshow criteria were used to describe performance of discrimination [16]. These criteria considered an AUROC value of 0.5 as no discrimination, 0.5 to <0.7 as poor discrimination, 0.7 to <0.8 as acceptable discrimination, 0.8 to <0.9 as excellent discrimination, and ≥0.9 as outstanding discrimination. Sensitivity, specificity, and negative predictive value (NPV) were defined using conventional definitions. Diagnostic yield was defined as the percentage of patients with a risk score below a given cutoff. All analyses were prespecified. The sample size of this analysis was driven by data availability in this multicenter database.

Analyses were performed in Python (Version 3.7.5; Python Software Foundation) using the XGBoost package (Version 0.82) [17] and the Scikit-Learn library (Version 0.21.3) [18]. The use of deidentified databases as described here met the non–human subjects research by the University of Vermont's Institutional Review Board criteria.

Results

Demographics and Proportion of PCR-Positive Patients in Training Data Set, External Validation Data Set, and Sensitivity Analysis Data Set

The training data set included 12,183 ED visits at 43 centers from the PHD, of which 2183 results were PCR-positive. The validation data set included 172,754 ED visits from 23 centers (21 from the PHD, as well as the independently collected data from CSMC and BIDMC), of which 1020 results were PCR-positive. The sensitivity analysis data set included 7842 records from 21 centers in the PHD holdout group. Patient demographics and visit characteristics are summarized in Table 1.

A total of 192,779 eligible patients were included in the study; the median age decile was 50 (IQR 30-60) years and 40.5% (78,249/192,779) were male. In the training, external validation, and sensitivity analysis data sets, the median age deciles were 50 (IQR 30-70) years, 50 (IQR 30-60) years, and 50 (IQR 40-70) years, respectively. Males represented 42.9%, 40.1%, and 47.4% of patients in the data sets, respectively.

Table 1. Demographics of patients and encounter details, by COVID-19 status[a].

Demographics	Training (N=12,183)		External validation (N=172,754)		Sensitivity analysis (N=7842)	
	Negative (n=10,000)	Positive (n=2183)	Negative (n=171,734)	Positive (n=1020)	Negative (n=6890)	Positive (n=952)
Age (years), n (%)						
20 to <30	1392 (14)	198 (9)	27,952 (16)	71 (7)	709 (10)	70 (7)
30 to <40	1481 (15)	304 (14)	29,187 (17)	127 (12)	882 (13)	119 (12)
40 to <50	1398 (14)	413 (19)	27,764 (16)	214 (21)	896 (13)	205 (22)
50 to <60	1649 (16)	400 (18)	28,896 (17)	217 (21)	1172 (17)	208 (22)
60 to <70	1512 (15)	367 (17)	23,771 (14)	180 (18)	1200 (17)	163 (17)
70 to <80	1322 (13)	264 (12)	18,460 (11)	121 (12)	1063 (15)	108 (11)
≥80	1246 (12)	237 (11)	15,704 (9)	90 (9)	968 (14)	79 (8)
Gender, n (%)						
Female	5876 (59)	1079 (49)	102,942 (60)	502 (49)	3650 (53)	477 (50)
Male	4122 (41)	1104 (51)	68,790 (40)	518 (51)	3240 (47)	475 (50)
Unknown	2 (0)	0 (0)	2 (0)	0 (0)	0 (0)	0 (0)
Race, n (%)						
Black	1791 (18)	397 (18)	28,874 (17)	212 (21)	1230 (18)	201 (21)
Other	904 (9)	976 (45)	23,222 (14)	453 (44)	772 (11)	448 (47)
Unknown	450 (4)	102 (5)	12,284 (7)	48 (5)	368 (5)	36 (4)
White	6855 (69)	708 (32)	107,354 (63)	307 (30)	4520 (66)	267 (28)
Census division[b], n (%)						
East North Central	2065 (21)	280 (13)	16,184 (9)	108 (11)	1103 (16)	108 (11)
East South Central	0 (0)	0 (0)	3549 (2)	50 (5)	138 (2)	50 (5)
Middle Atlantic	782 (8)	294 (13)	18,776 (11)	92 (9)	1356 (20)	92 (10)
New England	493 (5)	1 (0)	31,624 (18)	1 (0)	1 (0)	1 (0)
Pacific	106 (1)	32 (1)	3617 (2)	69 (7)	34 (0)	1 (0)
South Atlantic	3116 (31)	1192 (55)	70,463 (41)	613 (60)	2790 (40)	613 (64)
West North Central	633 (6)	39 (2)	0 (0)	0 (0)	0 (0)	0 (0)
West South Central	2805 (28)	345 (16)	27,521 (16)	87 (9)	1468 (21)	87 (9)
Rural or urban[b], n (%)						
Rural	583 (6)	21 (1)	3617 (2)	1 (0)	34 (0)	1 (0)
Urban	9417 (94)	2162 (99)	168,117 (98)	1019 (100)	6856 (100)	951 (100)
Disposition, n (%)						
Discharge from emergency department	7487 (75)	1175 (54)	132,195 (77)	522 (51)	4072 (59)	522 (55)
Non–intensive care unit admission	2068 (21)	805 (37)	29,793 (17)	379 (37)	2375 (34)	335 (35)
Intensive care unit admission	445 (4)	203 (9)	9746 (6)	119 (12)	443 (6)	95 (10)

[a] For the training data set: COVID-19 positivity was defined as a positive COVID-19 reverse-transcription polymerase chain reaction (hereafter, PCR) test on the day of presentation to the emergency department among patients in the pandemic time frame (March 2020 through July 2020) in the Premier Healthcare Database (PHD) database among a random selection of 43 of the 64 PHD hospitals reporting PCR positives. COVID-19 negativity was defined as a selection of 10,000 patients in the prepandemic time frame (January through December 2019) in the PHD database from the same 43 hospitals as the patients with COVID-19. For the external validation data set: COVID-19 positivity was defined the same as for the training data set for the PHD data set but also included 952 PCR-positives from the 21 hospitals in the PHD holdout set. Additionally, it included 68 patients with

PCR-confirmed COVID-19 from Cedar Sinai Medical Center from March and April 2020. COVID-19 negativity in the external validation set was defined using 154,341 prepandemic visits from the 21 hospitals in the PHD holdout set (January through December 2019) in which primary diagnoses were among the 20 most frequent primary diagnoses given to patients negative for COVID-19 during the pandemic, using Clinical Classification Software Refined codes. It also included 17,393 prepandemic (2008-2019) patient encounters from Beth Israel Deaconess Medical Center. For the sensitivity data set: COVID-19 positivity included the same 952 PCR-positives from the 21 hospitals in the external validation data set. COVID-19 negativity was defined as visits with at least 1 PCR-negative but no PCR-positive result on the day of presentation, and included all 6890 patients with such results from the same 21 hospitals as the positives.

[b] Census division was defined using US Census classification [19]. Rural areas are considered territory outside of the US Census Bureau's definition of urban [20]. These geographic descriptions pertain to the hospital, not the patient's permanent residence.

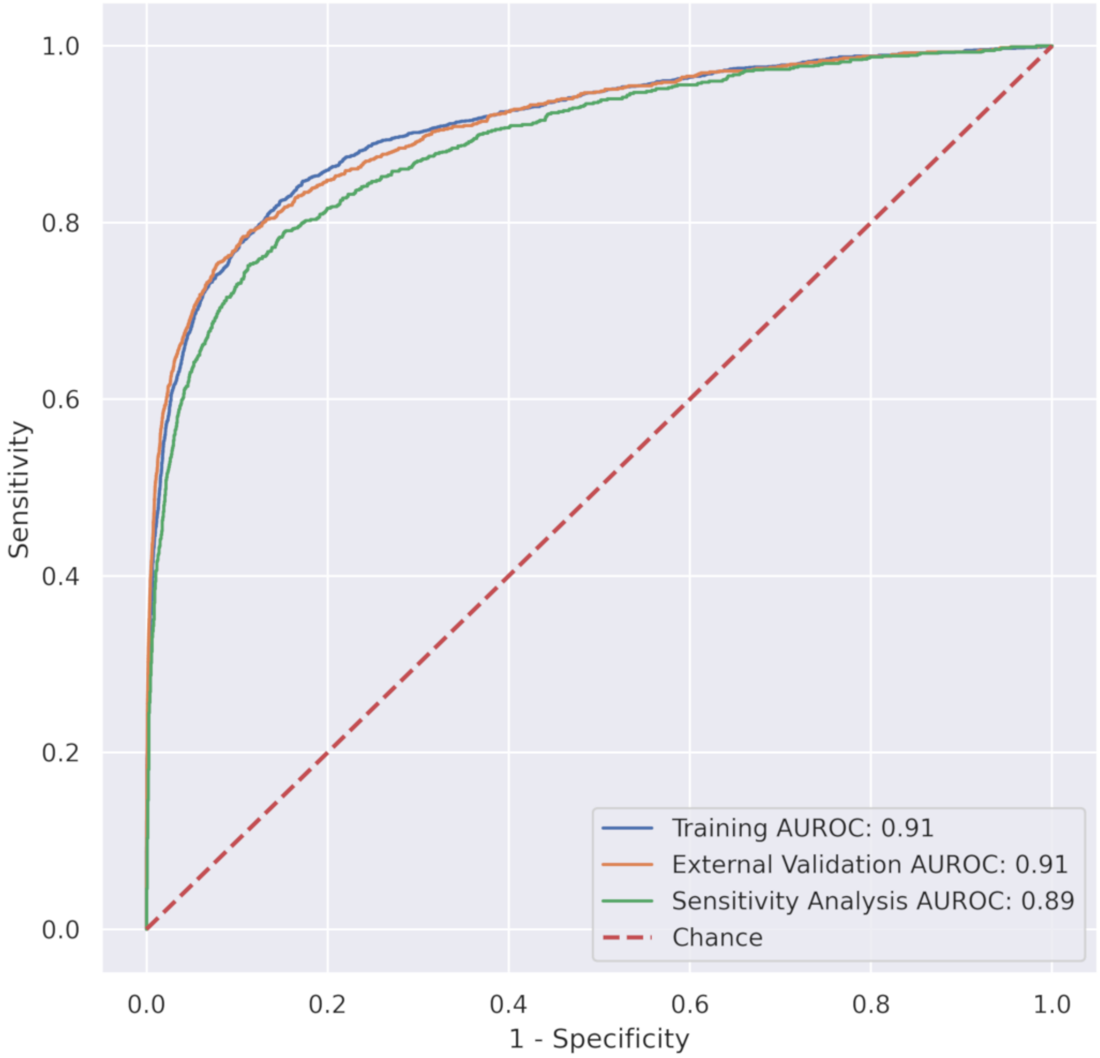
Selected Features Included in the Model and Individual Feature Performance

The RFECV method led to the final set of 15 features listed in Table S1 in Multimedia Appendix 1. The distributions of these features in the training data set, stratified by COVID-19–positive and COVID-19–negative status and ordered by importance to the model, are shown in Figure S1 in Multimedia Appendix 1. The features with the largest calculated importance were eosinophils, calcium, and aspartate aminotransferase. Summary statistics of these features in the training, external validation, and sensitivity analysis data sets, stratified by COVID-19 status, appear in Table S1 in Multimedia Appendix 1.

Performance of Individual Features and Model Performance in the Training Data Set

The AUROC for each individual feature in the training data set is shown in Figure S2 in Multimedia Appendix 1. The highest AUROCs were observed for eosinophils, calcium, and aspartate aminotransferase (0.70-0.80). The final model's AUROC in the training data set was 0.91 (95% CI 0.90-0.92; Figure 1).

Figure 1. Discrimination as assessed by ROC curves for training, external validation, and sensitivity analysis data sets. ROC curves for the 3 different data sets: training (blue), external validation (orange), and sensitivity analysis (green). The training curve was obtained through 5-fold cross-validation, where positive controls are PCR-confirmed cases during the pandemic (N=2183) and negative controls are prepandemic patients (N=10,000) from 43 hospitals in the PHD. The training AUROC was 0.91 (95% CI 0.90-0.92). The external validation curve was performed in the external validation data set after training the model on the training data set. External validation positives are PCR-confirmed cases from Cedars-Sinai Medical Center (N=68) and from the PHD holdout set (N=952) comprising 21 hospitals. External validation negatives are prepandemic (2019) patients, from the same 21 PHD hospitals, that match the top 20 primary non–COVID-19 diagnoses in 2020 (N=154,341), as well as all eligible prepandemic (2008-2019) Beth Israel Deaconess Medical Center patients (N=17,393). The AUROC in the external validation data set was 0.91 (95% CI 0.90-0.92). The sensitivity analysis curve demonstrates the effect of using prepandemic patients as negative controls compared to using PCR-negatives from 2020. In this data set, both positives (N=952) and negatives (N=6890) were PCR-confirmed patients from the PHD holdout set (21 hospitals), and no prepandemic data was included. The AUROC in the sensitivity analysis set was 0.89 (95% CI 0.88-0.90). AUROC: area under the receiver operating characteristic curve; PCR: polymerase chain reaction; PHD: Premier Healthcare Database; ROC: receiver operating characteristic.



Model Performance in the External Validation Data Set

The model’s AUROC in the external validation data set was 0.91 (95% CI 0.90-0.92), as shown in Figure 1. This corresponds to an outstanding discrimination per the Hosmer-Lemeshow criteria [16]. Sensitivity and specificity were 95.9 and 41.7 at a score cutoff of 1, 92.6 and 60.0 at a score of 2, 85.5 and 78.5 at a cutoff of 5, and 79.4 and 87.6 at a cutoff of 10, respectively (Table 2).

With a COVID-19 population prevalence of 1%, each of these cutoffs had an NPV >99%; at 10% prevalence, each was >97%, and at a prevalence of 20%, each was >94%. The diagnostic yield ranged from 34% (20% prevalence, score cutoff of 1) to 87% (1% prevalence, score cutoff of 10).

Table 2. Clinical performance metrics for the model in the external validation data set for various score cutoffs and COVID-19 pretest prevalence[a].

Score cutoff	Sensitivity	Specificity	Likelihood ratio[b]	Prevalence of 1%		Prevalence of 10%		Prevalence of 20%	
				NPV[c], %	Yield[d], %	NPV, %	Yield, %	NPV, %	Yield, %
1	95.9	41.7	0.099	99.9	41.3	98.9	38.0	97.6	34.2
2	92.6	60.0	0.124	99.9	59.4	98.6	54.7	97.0	49.5
5	85.5	78.5	0.185	99.8	77.8	98.0	72.1	95.6	65.7
10	79.4	87.6	0.235	99.8	86.9	97.4	80.9	94.4	74.2

[a] The maximum score was 100; a higher score indicates higher model prediction of COVID-19 positivity.

[b] The likelihood ratio uses the equation for negative tests.

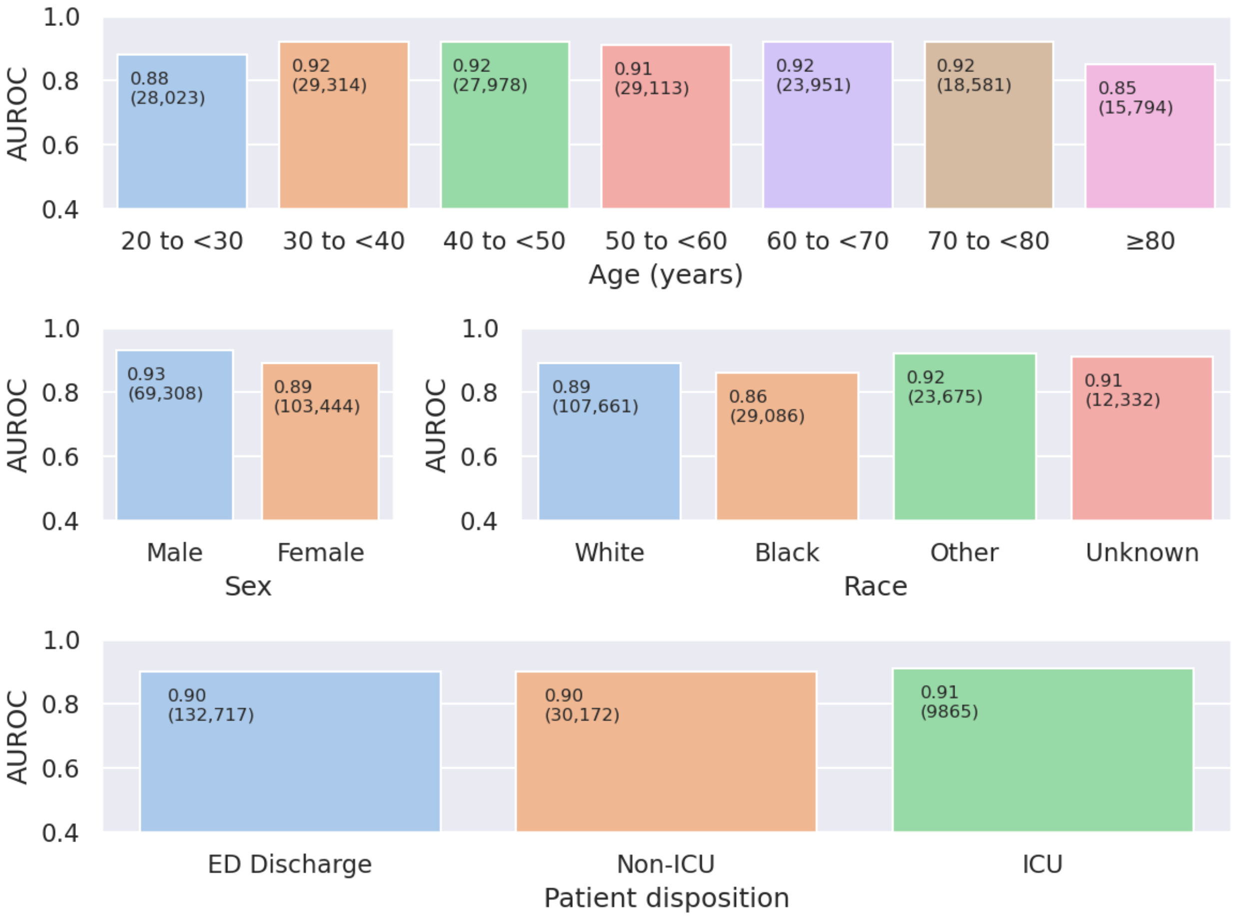
[c] NPV: negative predictive value.

[d] Yield refers to diagnostic yield, which is the percentage of patients that can be ruled out (ie, those with a score below the cutoff).

Sensitivity Analysis and Subgroup Analyses

Figure 1 depicts the ROC curve in the sensitivity analysis data set, which contains only year 2020 patients with PCR-confirmed positive and negative results (ie, no historical negatives). The AUROC was 0.89 (95% CI 0.88-0.90). In Figure 2, the AUROC is presented for various demographic cohorts as well as patient disposition (ED discharge, non-ICU, and ICU) in the external validation data set. AUROCs ranged from 0.86 to 0.93. AUROC by these subgroups was similar in the sensitivity analysis data set, which appears as Figure S3 in the Multimedia Appendix 1.

Figure 2. Discrimination as assessed by AUROC curve in age, sex, race, and ED disposition subgroups in the external validation data set. Non-ICU patients were admitted to the hospital but not to an ICU. Distribution of AUROC curves per demographic, as well as per patient disposition type (ED discharge, non-ICU, and ICU) in the external validation data set. Top numbers are AUROC curves, bottom numbers in parentheses are the number of patients. AUROC: area under the receiver operating characteristic; ED: emergency department; ICU: intensive care unit.



Distribution of Risk Scores in Selected Subgroups

Lastly, an extensive distribution of risk scores for various subgroups is shown in Figure S4 in Multimedia Appendix 1, including prepandemic patients whose primary diagnoses were among the top 20 primary diagnoses among patients without COVID-19 in 2020. From visual inspection, it can be seen that high scores track PCR-positive patients consistently across all cohorts.

Discussion

Principal Results

A development and external validation study of a machine learning model for COVID-19 status using laboratory tests routinely collected in adult ED patients found high discrimination across age, race, sex, and disease severity subgroups. This model had high diagnostic yield at low score cutoffs in a screening population with a disease prevalence of <10%. Such a model could rapidly identify those at low risk for COVID-19 in a “rule-out” method, and might reduce the need for PCR testing in such patients.

Comparison With Prior Work

Prior literature has described the application of machine learning techniques to commonly collected laboratory data for estimation of missing laboratory analytes. For example, an analysis by Waljee and colleagues [21] leveraged machine learning techniques for imputation of missing laboratory data in cohorts of patients with cirrhosis and inflammatory bowel disease at a single institution. In comparison to other common imputation techniques described in this manuscript, the machine learning technique introduced the least imputation error for these laboratory data. Luo and colleagues [22] used similar methods to estimate ferritin from a single medical center, and found that the machine learning technique outperformed traditional imputation methods. These serve as strong evidence of the potential use of machine learning for use in estimation of laboratory data. However, outside of imputation of missing values from research databases, the clinical utility for such techniques was unclear prior to the COVID-19 pandemic.

During this pandemic, there is an urgent need to rapidly identify patients with the disease to inform supportive clinical care. Prior work has attempted to integrate combinations of clinical data points in diagnostic models, though only a few are currently published in peer-reviewed literature [23]. The selection of the specific data points to integrate into machine learning models for COVID-19 diagnosis has implications on integration into existing clinical delivery. In contrast with the results here, which only included components of the routinely collected complete blood count with differential and complete metabolic panel laboratory tests, others have integrated nonlaboratory features. Sun and colleagues [24] reported 3 models including demographics, radiological data, and symptomatology, and obtained AUCs ranging from 0.65 to 0.88 for these models. Symptomatology was not obtained with structured, validated questionnaires and the ability to capture these symptoms in a reproducible manner might be difficult outside of a research setting. Further, modern medical records cannot integrate such symptoms into automated risk scores as they are not documented in a structured way.

Structured data obtained routinely in clinical examinations are the simplest to integrate, and might have the least variability between institutions. These include vital signs, demographics, laboratory findings, and radiological images. There are few studies describing the use of such data for the diagnosis of COVID-19. One study found a machine learning method had an accuracy of 87% for distinguishing between COVID-19 from pneumonia or no relevant findings using chest radiographs [25]. A different model developed from chest computed tomography images reported an AUROC of 0.994 when distinguishing between COVID-19 and atypical or viral pneumonia [26]. However, national organizations recommend against the use of radiological imaging for diagnosis of COVID-19, in part because of the added risk of spreading infection through additional visitation to radiology suites [27]. These models are unlikely to be readily deployed because of the challenges of performing elective radiological tests during this pandemic.

An additional consideration in the development of machine learning models is the inclusion of an adequate sample size for model training [28-30]. Other studies have investigated the role of laboratory data with or without other nonradiological structured clinical data or demographics for the diagnosis of COVID-19 using machine learning techniques. For example, Wu and colleagues [31] reported a C-index of 0.99 but included only 108 patients (12 COVID-19–positive) in their training. Similarly, individual efforts led by Batista, Brinati, and Soares [32-34] describe machine learning models trained on 234, 279, and 599 patients, respectively. These studies are also limited in the small number of centers from which patients were enrolled, and lack of diversity in their patient populations.

Advancement of Scientific Knowledge

The present analysis advances science in several key ways. First, we describe a machine learning model developed in a diverse patient population with routine laboratory data from multiple clinical centers across the United States [35]. Second, the model incorporates common laboratory tests that are widely available with rapid turnaround time. As the machine learning model can be performed essentially instantaneously, the primary time limitation is related to phlebotomy and specimen processing. There is a well-known bottleneck in completing conventional COVID-19 PCR assays; a commercial laboratory recently reported a 7-day reporting lag [36]. Third, the present model could identify those at lowest risk for COVID-19 to inform a “rule-out” method for screening. Those with intermediate or greater risk for COVID-19 could be further assessed with COVID-19 PCR testing, if indicated. Depending on the selected score cutoff and population prevalence, such an approach could rule out 34% to 87% of ED patients requiring conventional COVID-19 PCR testing (see Yield, Table 2). The specific score cutoff for rule out of COVID-19 with this model can be customized based upon what an institution considers to be an “acceptable” target NPV. However, the diagnostic yield will change based upon the screening population prevalence of COVID-19, and the diagnostic yield will be inversely related to the screening population prevalence of COVID-19. For example, assume that an institution determines that an acceptable NPV for this model is 97.5%. If this institution's screening population has a 20% prevalence of COVID-19, the threshold score cutoff would be set at 1, and the diagnostic yield (ie, the percentage of patients ruled out for COVID-19 at a score cutoff of 1) would be 34.2% (Table 2). However, at a prevalence of 10%, the score cutoff threshold would be 10, and the diagnostic yield would be 80.9%. The efficiency of diagnostic yield with this model is higher at lower prevalence. Finally, the sensitivity of the present model at a score cutoff of 1, 2, and 5 (95.9, 92.6, and 85.5, respectively) was similar to COVID-19 antigen assays (66.1-86.3) and sputum and saliva PCR assays (62.3-97.2) [11]. The comparatively similar sensitivities between the model and these existing assays supports the clinical utility of machine learning models as future diagnostic tools.

Weaknesses and Strengths

This study has weaknesses. Although the choice of prepandemic controls partially circumvents the issue of false negatives in PCR testing by ensuring the negatives that the model is trained on are true negatives, it does not ensure that the positives encompass the full spectrum of true positives, since those are sometimes missed by PCR due to changes in viral load as a function of disease progression [37,38]. Additionally, the use of controls from a different time period could introduce a bias of its own, such as different demographics or non–COVID-19 morbidities. However, the sensitivity analysis used COVID-19 positives and negatives from the pandemic time frame, and the performance of the model was reassuringly similar to the performance in the external validation. The performance in demographic, clinical diagnosis, and ED disposition subgroups was also similar to the external validation. Laboratory data were performed locally at each hospital, rather than centrally. The model requires all components of the laboratory data to be included. This study only included patients who visited an ED. Although it is likely that some of the patients in this study were asymptomatic or presymptomatic and were found to have COVID-19 as part of routine admission, we were unable to determine the indications for screening and therefore are unable to determine the performance of this model in asymptomatic and presymptomatic adults. The present analysis only accounted for results from COVID-19 PCR tests and not for alternative diagnostic methods, such as antigen testing for acute infection or antibody testing to demonstrate prior infection. Finally, the research database did not include details about the specific PCR assay used in diagnosis, so we are unable to comment on performance of the model in comparison to the performance of the specific assays.

This study has strengths. This study included data from a large number of patients and hospitals, and to our knowledge is the largest application of machine learning to COVID-19. Data were derived from an electronic medical records database that is commonly used in clinical research. The patient population was geographically and racially diverse. The only features included in the model are those included in blood tests that are already routinely collected in ED encounters. Further, these tests were from multiple hospitals, suggesting that the model is robust against different specimen collection, handling practices, and instrumentation. Sensitivity analyses were performed to evaluate potential biases due to the choice of prepandemic negative controls, and no significant bias was observed across multiple dimensions. Our methods extend on established machine learning–based imputation methods for missing laboratory data [21,22], and suggest there may be clinical utility of these techniques in ruling out the disease. Finally, the external validation was a true external validation since it used data from hospitals that were not included in the training data set. This supports the resilience of the model across institutions with differing specimen handling and laboratory processing methods.

Conclusions

A machine learning model for ruling out COVID-19 in ED patients that integrates commonly collected laboratory data had a discrimination accuracy that can be classified as excellent to outstanding [16]. Using score cutoffs of 5 and 10 points, and assuming a 10% screening population prevalence of COVID-19, 72% and 81% of patients were ruled out with this model while maintaining an NPV >97%, respectively.

Acknowledgments

We thank Alistair EW Johnson, PhD, and John House, MS, for providing technical expertise related to the BIDMC and PHD data sets, respectively. We would also like to thank Ryan McGinnis, PhD, who provided helpful feedback on the manuscript, and Dr Robert A Levine, MD; R George Hauser, MD; and Mark K Fung, MD, for various answers to laboratory medicine questions. We thank Susan Gifford for her helpful review of this manuscript.

Conflicts of Interest

TBP, AB, ASW, and ICJ have no disclosures. VFT received research funding from the National Institutes of Health (paid to the institution) for COVID-19–related clinical trials. ANB is a paid intern at Biocogniv. TSK and ABA have ownership of Biocogniv.

Multimedia Appendix 1
Supplementary figures and tables.
[DOCX File , 1012 KB - jmir_v22i12e24048_app1.docx]

References

1. Covid in the U.S.: Latest Map and Case Count. The New York Times. URL: https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html [accessed 2020-11-23]
2. Baden LR, Rubin EJ. Covid-19 — The Search for Effective Therapy. N Engl J Med 2020 May 07;382(19):1851-1852. [doi: 10.1056/nejme2005477]
3. Karow J. Scientists seek solution to coronavirus testing bottleneck. Modern Healthcare. URL: https://www.modernhealthcare.com/patients/scientists-seek-solution-coronavirus-testing-bottleneck [accessed 2020-08-04]
4. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J 2017 Jun 14;38(23):1805-1814 [FREE Full text] [doi: 10.1093/eurheartj/ehw302] [Medline: 27436868]
5. Peterson ED. Machine Learning, Predictive Analytics, and Clinical Practice: Can the Past Inform the Present? JAMA 2019 Nov 22:1. [doi: 10.1001/jama.2019.17831] [Medline: 31755902]
6. Waljee AK, Higgins PDR. Machine learning in medicine: a primer for physicians. Am J Gastroenterol 2010 Jun;105(6):1224-1226. [doi: 10.1038/ajg.2010.173] [Medline: 20523307]
7. Kearon C, de Wit K, Parpia S, Schulman S, Afilalo M, Hirsch A, PEGeD Study Investigators. Diagnosis of Pulmonary Embolism with d-Dimer Adjusted to Clinical Probability. N Engl J Med 2019 Nov 28;381(22):2125-2134. [doi: 10.1056/NEJMoa1909159] [Medline: 31774957]
8. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. Clin Chem 2015 Dec;61(12):1446-1452. [doi: 10.1373/clinchem.2015.246280] [Medline: 26510957]
9. Premier Healthcare Database (COVID-19) white paper: Data that informs and performs. Premier Inc. 2020 Mar 02. URL: https://products.premierinc.com/downloads/PremierHealthcareDatabaseWhitepaper.pdf [accessed 2020-11-23]
10. Centers for Disease Control and Prevention. COVID-19 Stats: COVID-19 Incidence,* by Urban-Rural Classification - United States, January 22-October 31, 2020. MMWR Morb Mortal Wkly Rep 2020 Nov 20;69(46):1753 [FREE Full text] [doi: 10.15585/mmwr.mm6946a6] [Medline: 33211682]
11. Böger B, Fachi MM, Vilhena RO, Cobre AF, Tonin FS, Pontarolo R. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. Am J Infect Control 2020 Jul 10:1 [FREE Full text] [doi: 10.1016/j.ajic.2020.07.011] [Medline: 32659413]
12. Axell-House DB, Lavingia R, Rafferty M, Clark E, Amirian ES, Chiao EY. The estimation of diagnostic accuracy of tests for COVID-19: A scoping review. J Infect 2020 Nov;81(5):681-697 [FREE Full text] [doi: 10.1016/j.jinf.2020.08.043] [Medline: 32882315]
13. Redman T. If your data is bad, your machine learning tools are useless. Harvard Business Review. URL: https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless [accessed 2020-08-05]
14. sklearn.feature_selection.RFECV. scikit-learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFECV.html [accessed 2020-08-01]
15. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16; August 2016; San Francisco, CA, USA p. 785-794. [doi: 10.1145/2939672.2939785]
16. Hosmer JD, Lemeshow S, Sturdivant R. Applied Logistic Regression. Hoboken, NJ, USA: John Wiley & Sons Inc; 2013.
17. XGBoost GPU Support. XGBoost. URL: https://xgboost.readthedocs.io/en/release_0.82/gpu/index.html [accessed 2020-08-04]
18. Albon C. Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning. Sebastopol, CA, USA: O'Reilly Media; 2018.
19. 2010 census regions and divisions of the United States. United States Census Bureau. URL: https://www.census.gov/geographies/reference-maps/2010/geo/2010-census-regions-and-divisions-of-the-united-states.html [accessed 2020-08-04]
20. 2010 census urban and rural classification and urban area criteria. United States Census Bureau. URL: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html [accessed 2020-08-04]
21. Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, et al. Comparison of imputation methods for missing laboratory data in medicine. BMJ Open 2013 Aug 01;3(8):e002847 [FREE Full text] [doi: 10.1136/bmjopen-2013-002847] [Medline: 23906948]
22. Luo Y, Szolovits P, Dighe AS, Baron JM. Using Machine Learning to Predict Laboratory Test Results. Am J Clin Pathol 2016 Jun 21;145(6):778-788. [doi: 10.1093/ajcp/aqw064] [Medline: 27329638]
23. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
24. Sun Y, Koh V, Marimuthu K, Ng OT, Young B, Vasoo S, National Centre for Infectious Diseases COVID-19 Outbreak Research Team. Epidemiological and Clinical Predictors of COVID-19. Clin Infect Dis 2020 Jul 28;71(15):786-792. [doi: 10.1093/cid/ciaa322] [Medline: 32211755]
25. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 2020 Jun;121:103792 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103792] [Medline: 32568675]

26. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comput Biol Med 2020 Jun;121:103795 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103795] [Medline: 32568676]
27. ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection. American College of Radiology. URL: https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection [accessed 2020-08-05]
28. Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods 2015 Jul 30;250:126-136. [doi: 10.1016/j.jneumeth.2015.01.010] [Medline: 25596422]
29. Raudys S, Jain A. Small sample size effects in statistical pattern recognition: recommendations for practitioners. IEEE Trans Pattern Anal Machine Intell 1991;13(3):252-264. [doi: 10.1109/34.75512]
30. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One 2019 Nov 7;14(11):e0224365 [FREE Full text] [doi: 10.1371/journal.pone.0224365] [Medline: 31697686]
31. Wu J, Zhang P, Zhang L, Meng W, Li J, Tong C, et al. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. medRxiv. Preprint published online on April 6, 2020. [doi: 10.1101/2020.04.02.20051136]
32. Batista A, Miraglia J, Donato T, Chiavegatto FA. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. medRxiv. Preprint published online on April 14, 2020. [doi: 10.1101/2020.04.04.20052092]
33. Brinati D, Campagner A, Ferrari D, Locatelli M, Banfi G, Cabitza F. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. J Med Syst 2020 Jul 01;44(8):135 [FREE Full text] [doi: 10.1007/s10916-020-01597-4] [Medline: 32607737]
34. Soares F, Villavicencio A, Anzanello M, Fogliatto F, Idiart M, Stevenson M. A novel specific artificial intelligence-based method to identify COVID-19 cases using simple blood exams. medRxiv. Preprint published online on July 3, 2020 [FREE Full text] [doi: 10.1101/2020.04.10.20061036]
35. Rehmani R, Amanullah S. Analysis of blood tests in the emergency department of a tertiary care hospital. Postgrad Med J 1999 Nov 01;75(889):662-666 [FREE Full text] [doi: 10.1136/pgmj.75.889.662] [Medline: 10621876]
36. Picchi A. Quest Diagnostics says COVID-19 test turnaround now 7 days due to surge. CBS News. URL: https://www.cbsnews.com/news/quest-diagnostics-coronavirus-testing-seven-day-turnaround-case-surge/ [accessed 2020-08-03]
37. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure. Ann Intern Med 2020 Aug 18;173(4):262-267 [FREE Full text] [doi: 10.7326/M20-1495] [Medline: 32422057]
38. Woloshin S, Patel N, Kesselheim AS. False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications. N Engl J Med 2020 Aug 06;383(6):e38. [doi: 10.1056/nejmp2015897]

Abbreviations

AUC: area under the curve
AUROC: area under the receiver operating characteristic
BIDMC: Beth Israel Deaconess Medical Center
CCSR: Clinical Classifications Software Refined
CSMC: Cedars-Sinai Medical Center
ED: emergency department
ICU: intensive care unit
NPV: negative predictive value
PCR: reverse transcription polymerase chain reaction
PHD: Premier Healthcare Database
RFECV: recursive feature elimination with cross-validation
ROC: receiver operating characteristic
STARD: Standards for Reporting Diagnostic Accuracy Studies

Edited by G Eysenbach; submitted 01.09.20; peer-reviewed by L Post, S Case, G Kannan; comments to author 23.09.20; revised version received 14.10.20; accepted 19.11.20; published 02.12.20.

Please cite as:
Plante TB, Blau AM, Berg AN, Weinberg AS, Jun IC, Tapson VF, Kanigan TS, Adib AB
Development and External Validation of a Machine Learning Tool to Rule Out COVID-19 Among Adults in the Emergency Department Using Routine Blood Tests: A Large, Multicenter, Real-World Study
J Med Internet Res 2020;22(12):e24048
URL: https://www.jmir.org/2020/12/e24048
doi:10.2196/24048
PMID:33226957

Original Paper

Mental Health Burden in Different Professions During the Final Stage of the COVID-19 Lockdown in China: Cross-sectional Survey Study

Junfeng Du[1,2*], MSc; Gwendolyn Mayer[2*], DipPsych; Svenja Hummel[2]; Neele Oetjen[2], MSc; Nadine Gronewold[2], MSc; Ali Zafar[2], MA, MSc; Jobst-Hendrik Schultz[2], PD, Dr med

[1]Liyuan Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of General Internal Medicine and Psychosomatics, Heidelberg University Hospital, Heidelberg, Germany

[*]these authors contributed equally

Corresponding Author:
Gwendolyn Mayer, DipPsych
Department of General Internal Medicine and Psychosomatics
Heidelberg University Hospital
Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 62215635685
Fax: 49 6221567505
Email: gwendolyn.mayer@med.uni-heidelberg.de

Abstract

Background: COVID-19 resulted in considerable mental health burden in the Chinese general population and among health care workers at the beginning and peak of the pandemic. However, little is known about potentially vulnerable groups during the final stage of the lockdown.

Objective: The aim of this survey study was to assess the mental health burden of different professions in China in order to find vulnerable groups, possible influencing factors, and successful ways of coping during the last 4 weeks of the lockdown in Hubei Province.

Methods: A cross-sectional online survey asked participants about current residence, daily working hours, exposure to COVID-19 at work, and media preferences. We used a shortened version of the Depression, Anxiety and Stress Scale (DASS-21) to assess mental health. Further assessments included perceived stress (Simplified Chinese version of the 14-item Perceived Stress Scale), coping strategies for all participants, and specific stressors for health care workers. We followed the reporting guidelines of the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement for observational studies.

Results: The sample (N=687) consisted of 158 doctors, 221 nurses, 24 other medical staff, 43 students, 60 teachers/government staff, 135 economy staff, 26 workers/farmers, and 20 professions designated under the “other” category. We found increased depression (n=123, 17.9%), anxiety (n=208, 30.3%), and stress (n=94, 13.7%) in our sample. Professions that were vulnerable to depression were other medical staff and students. Doctors, nurses, and students were vulnerable to anxiety; and other medical staff, students, and economy staff were vulnerable to stress. Coping strategies were reduced to three factors: active, mental, and emotional. Being female and emotional coping were independently associated with depression, anxiety, or stress. Applying active coping strategies showed lower odds for anxiety while mental coping strategies showed lower odds for depression, anxiety, and stress. Age, being inside a lockdown area, exposure to COVID-19 at work, and having a high workload (8-12 hours per day) were not associated with depression, anxiety, or stress. WeChat was the preferred way of staying informed across all groups.

Conclusions: By the end of the lockdown, a considerable part of the Chinese population showed increased levels of depression and anxiety. Students and other medical staff were the most affected, while economy staff were highly stressed. Doctors and nurses need support regarding potential anxiety disorders. Future work should focus on longitudinal results of the pandemic and develop targeted preventive measures.

(J Med Internet Res 2020;22(12):e24240) doi:10.2196/24240

KEYWORDS

mental health; COVID-19; China; depression; anxiety; lockdown; coping strategies; stressors; stress; doctors; nurses; students; media consumption; WeChat

Introduction

Background

In December 2019, pneumonia cases of unknown etiology in Wuhan, Hubei Province, were reported by Chinese authorities. By January 3, 2020, 44 cases requiring hospitalization were officially confirmed [1]. The disease was titled COVID-19 by the World Health Organization in February 2020 [2]. As the infection spread rapidly all over the country [3], the public health response of the government included the largest quarantine in China's history. Starting from January 23, the lockdown in Hubei Province, the epicenter of COVID-19, lasted 76 days and ended on April 8, 2020 [4]. Zhejiang was another province that was put under lockdown [5].

The pandemic put a considerable psychological burden on citizens, which was not simply due to fear of infection but also isolation, helplessness, and grief over the loss of relatives without having the opportunity to take leave or to organize a funeral. Even more aggravating was that trusted persons, like family and friends, could be infected, and thus, became part of an invisible danger [6].

Previous epidemics, like the severe acute respiratory syndrome (SARS) in Hong Kong in 2003 or the Middle East respiratory syndrome (MERS) in Saudi Arabia in 2012, have taught us to care for the mental health of the general population and frontline health care workers [7-10]. In Hong Kong, suicide rates among older adults increased significantly in 2003 and 2004 [11]. It is noteworthy to recall that high-risk health workers suffered from long-lasting depression and posttraumatic stress responses even 1 year after the SARS outbreak [12].

Mental Health and Coping Strategies

Several studies have investigated the mental health consequences of the ongoing pandemic in the Chinese population and its strategies to successfully cope with the demanding situation. Wang et al [13] found increased anxiety in nearly 30% of 1210 participants, especially caused by worrying about family members. The same authors also reported longitudinal results, repeating the measures after 4 weeks in March 2020 [14], and found no changes in the scores despite increased infection rates. A high level of trust in doctors and health information, as well as personal protective measures, was rated as helpful. The harmful effects of hoarding food and medicine were described as being associated with elevated levels of depression at the beginning of the crisis [15]. Another study found that cognitive coping behavior (eg, obtaining knowledge about COVID-19) and prosocial coping styles (eg, adherence to social distancing) proved to be protective for the population [16]. These authors described the high impact of the pandemic on the livelihood of the population and examined the harmful effects of media exposure on mental health.

Besides the obvious impact of the pandemic on mental health like the fear of infection and isolation due to quarantine measures [17], excessive media consumption was linked to mood disorders during the lockdown in China [18,19]. However, little is known about the influence of media preferences on the mental health of the population during the crisis.

The Mental Health of Vulnerable Groups During COVID-19

Since health care workers at the frontline were exposed to particularly demanding conditions during the peak of the pandemic, their mental health and coping strategies have become an early issue of concern. One of the first studies on this topic focused on medical and nursing staff in Wuhan and found elevated levels of subthreshold mental health disturbances in nearly 40% of the 994 participants surveyed [20]. A nationwide study in February reported nearly 5% of medical and nonmedical staff with moderate and high levels of anxiety and about 13% with depression [21]. In this study, nurses and young personnel were found to be particularly at risk for mental distress. A nationwide survey by the end of February showed even higher levels (anxiety: 13%; depression: 12%; insomnia: 38%) for doctors and nurses compared to nonmedical health workers [22]. Risk factors included living in a rural area, being female, and exposure to COVID-19. Additionally, health care workers were burdened by specific clinical and nonclinical stressors (eg, fear of bringing the virus home to family members and the experience of losing colleagues) [23].

There were some specific results on the psychological burden felt by nurses. Nurses in Anhui showed strong emotional responses. Increased exposure to COVID-19 cases evoked more anxiety and anger [24]. Increased levels of insomnia were reported among nurses in Wuhan, which might have been caused by comorbid sleep apnea due to stress [25]. Despite showing symptoms of severe distress, these Wuhan nurses refused to accept psychological help at the beginning of the pandemic [26].

Another vulnerable group included students, the majority of whom lived in quarantine with their families and reported victimization by facing or witnessing various stressful events related to COVID-19 [27]. Other studies found that the COVID-19 crisis impacted sleep quality [28], and increased anxiety among students was reported even after the lockdown ended [29]. Teachers were also affected not only by the outbreak but also by the stress experienced by their pupils [30].

In spite of the many studies regarding the mental health of the general population and health care workers on the frontline of the pandemic, we found no data on further vulnerable groups and professions that may be mentally or emotionally affected by indirect means. Although Huang and Zhao [31] found comparable depressive symptoms among employees in enterprises as in health care workers, and Wang et al [13,14] reported longitudinal results in February and March, these results were from the initial stages of the lockdown and data from the end of the lockdown are missing to date.

Objectives of the Study

The aim of this survey study was to assess the psychological burden of COVID-19 on the mental health of the Chinese population during the last 4 weeks of the lockdown in Hubei Province. We examined different professions in order to find vulnerable groups, possible influencing factors, and successful ways of coping. Moreover, we looked for specific stressors among doctors and nurses.

Methods

Study Design

We used a cross-sectional online survey design in order to investigate the impact of the COVID-19 pandemic on the mental health, stress, specific stressors, and coping strategies of different groups of the Chinese population. The study team of Heidelberg University Hospital developed the concept and the questionnaire, which was translated into Chinese. Its implementation into an online format and sampling was carried out by a publicity enterprise in Wuhan. The Tongji Medical College of the Huazhong University of Science and Technology supported the study by disseminating the link. The study started on March 19 and data were included until April 7. The lockdown in Hubei was officially ended on April 8 by the government [4].

Ethical approval for this study was granted by the Ethics Commission of the Medical Faculty of Heidelberg (S-361/2020). We followed the reporting guidelines of the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement for observational studies [32].

Measures

The questionnaire was derived from validated instruments and structured into four major parts. The first part asked for demographic data (place of residence, gender, age, marital status, educational background, and occupation), exposure to people infected with COVID-19 in general and at work, working hours per day, and media platforms used to obtain information (multiple choice). The second part asked for mental health parameters like depression, anxiety, and stress, measured by a shortened version of the 21-item Depression, Anxiety and Stress Scale (DASS-21) using a 4-point Likert scale [33]. We used the validated Chinese translation [34]. The instrument refers to a time span of the past week and has been shown to distinguish well between symptoms of depression, anxiety, and stress in clinical and nonclinical samples [35]. In order to assess stress levels during the past month, we used the 14-item Perceived Stress Scale (PSS-14) that explicitly refers to a longer time span using a 5-point Likert scale [36]. Again, we used a validated translation (ie, CPSS-14) [37]. The third part of our questionnaire targeted health care workers only and consisted of a questionnaire that was used in a former study on the SARS outbreak in 2003 [38]. The items asked for specific disease-related stressors of doctors and nurses and were rated on a 4-point Likert scale (0=not at all; 1=slightly; 2=moderately; 3=very much). Finally, the fourth part was again available for all participants and asked for successful coping strategies. The items were again taken from the SARS study [38] and recorded the frequency of use of various coping strategies on a 4-point scale (0=almost never; 1=sometimes; 2=often; 3=almost always). Both scales were translated by a Chinese native speaker (JD).

Data Analysis

The responses of the participants were downloaded from the online survey tool and further processed and analyzed using SPSS 24 (IBM Corp) [39]. We collected 1006 data sets and removed all data sets that were filled out after April 7 (n=74) and all questionnaires filled out in less than 513 seconds (n=226), which corresponded to the lowest percentile of the mean processing time for all samples. In these cases, we assumed a lack of credibility if a participant took less than this amount of time. Finally, we removed all participants who were younger than the legal age (n=7) and who were not in China during the survey (n=9). We calculated descriptive statistics and reported frequencies, means, standard deviations, and percentages.

Participants answering from Hubei and Zhejiang provinces were regarded as being affected by the lockdown (n=460). All other participants were not directly affected by the lockdown (n=226).

The scoring of the DASS-21 are calculated as sum scores that have to be multiplied by two. The total depression subscale score was divided into normal (0-9), mild (10-13), moderate (14-20), severe (21-27), and extremely severe depression (28+). The anxiety subscale score was divided into normal (0-7), mild (8-9), moderate (10-14), severe (15-19), and extremely severe anxiety (20+). The total stress subscale score was divided into normal (0-14), mild (15-18), moderate (19-25), severe (26-33), and extremely severe stress (34+). Next, we grouped the levels of severity into normal–mild and moderate–extremely severe for each score. We decided to put mild symptoms into one group together with the normal level, since we considered mild symptoms of depression and anxiety to exist regardless of the pandemic [40].

The CPSS-14 scores were calculated by sum scores as well. We reported the CPSS-14 scores and DASS-21 scores nationwide for each profession. For deeper analysis we calculated Pearson correlations in order to assess the relationship of perceived stress during the past 4 weeks and mental health scores for depression, anxiety, and stress during the past week.

The coping strategies and major stressors were calculated as means and standard deviations. We carried out a factor analysis (principal component analysis [PCA] with varimax rotation) for all coping strategies. The Kaiser-Meyer-Olkin (KMO) and Bartlett Test indicated a sufficient cohesion of the variables (KMO=0.76) [41]. Finally, binary logistic regression models were calculated and thereby investigated the associations of gender, lockdown area, contact with COVID-19 infections at work, and coping strategies (factors) with the odds of belonging to moderate–extremely severe depression, anxiety, or stress group. A nonsignificant value of *P*=.40 in the Hosmer-Lemeshow test indicated the goodness of fit of the models [42], and a Nagelkerkes R^2 of 0.17 showed an acceptable coefficient of determination [41].

In all analyses, *P* values <.05 were considered statistically significant.

Results

Participants

The sample included 687 participants, 72.3% (n=496) of whom were female and 27.7% were (n=190) male. The mean age was 36.92 years (SD 9.83) with a range of 18-71 years. The participants consisted of doctors (n=158, 23.0%), nurses (n=221, 32.2%), other medical staff (n=24, 3.5%), students (n=43, 6.6%), teachers/government staff (n=60, 8.7%), economy staff (n=135, 19.7%), workers/farmers (n=26, 3.8%), and others (n=20, 2.9%). We combined doctors and dentists into one category. Other medical staff referred to health care professionals who were not doctors or nurses. Economy staff consisted of employees and self-employed individuals in the IT (information technology) and finance sectors.

A majority of the participants were from Hubei Province (n=449, 65.4%); 30 (4.4%) came from Jiangsu and 21 (3.1%) each from Shanxi and Guangdong. A small group (n=11, 1.6%) came from Zhejiang, which was affected by a lockdown like Hubei. Demographic characteristics and details of each professional group are summarized in Table 1.

Table 1. Demographic characteristics of the study participants.

Characteristics	Participants, n (%)
Age (years), mean (SD)	36.92 (9.83)
Gender	
Male	190 (27.7)
Female	496 (72.3)
Family status	
Single	146 (21.3)
Married	501 (72.9)
Divorced	30 (4.4)
Widowed	2 (0.3)
In a relationship	8 (1.2)
Children	
Yes	499 (72.6)
No	188 (27.4)
Level of education	
Middle school	10 (1.5)
High school	25 (3.6)
Junior college	168 (24.5)
Bachelor	384 (55.9)
Master	77 (11.2)
Doctorate	23 (3.3)
Profession	
Doctors/dentists	158 (23.0)
Nurses	221 (32.2)
Other medical staff (eg, volunteers, pharmacists, midwives)	24 (3.5)
Students	43 (6.3)
Teachers/government staff	60 (8.7)
Economy (eg, employees, self-employed, salespersons)	135 (19.7)
Workers/farmers	26 (3.8)
Others (eg, housewives)	20 (2.9)
Residence	
Hubei	449 (65.4)
Jiangsu	30 (4.4)
Guangdong	21 (3.1)
Shanxi	21 (3.1)
Shandong	17 (2.5)
Fujian	16 (2.3)
Sichuan	15 (2.2)
Shanghai	15 (2.2)
Hunan	14 (2.0)
Zhejiang	11 (1.6)
Provinces with less than 10 participants	78 (11.2)
Total	687 (100)

Perceived Stress and Mental Health

Perceived stress was measured with a mean score of 23.70 (SD 7.52). The mean values for DASS-21 depression was 6.62 (SD 7.80), for DASS-21 anxiety was 7.01 (SD 7.00), and for DASS-21 stress was 10.18 (SD 8.63). Perceived stress was significantly correlated with DASS-21 depression (r=0.61, P<.001), DASS-21 anxiety (r=0.57, P<.001), and DASS-21 stress (r=0.66, P<.001).

Findings on mental health status for each profession are reported in Table 2; the DASS-21 scores were put into categories normal–mild and moderate–extremely severe.

Table 2. Results of the Simplified Chinese version of the 14-item Perceived Stress Scale (CPSS-14) and the Depression, Anxiety and Stress Scale - 21 Items (DASS-21).

Profession	Participants, n	CPSS-14		DASS-21 depression		DASS-21 anxiety	DASS-21 stress	
		Mean (SD)	NM[a], n (%)	MES[b], n (%)	NM, n (%)	MES, n (%)	NM, n (%)	MES, n (%)
Doctors	158	23.16 (7.26)	134 (84.8)	24 (15.2)	106 (67.1)	52 (32.9)	138 (87.3)	20 (12.5)
Nurses	221	23.62 (7.19)	183 (82.8)	38 (17.2)	152 (68.8)	69 (31.2)	197 (89.1)	24 (10.9)
Other medical staff	24	22.25 (8.09)	19 (79.2)	5 (20.8)	17 (70.8)	7 (29.2)	20 (83.3)	4 (16.7)
Students	43	26.30 (7.79)	33 (76.7)	10 (23.3)	25 (58.1)	18 (41.9)	34 (79.1)	9 (20.9)
Teachers/ govt staff	60	22.98 (6.09)	51 (85.0)	9 (15.0)	44 (73.3)	16 (26.7)	53 (88.3)	7 (11.7)
Economy staff	135	23.93 (8.68)	108 (79.4)	27 (20.0)	102 (75.6)	33 (24.4)	112 (83.0)	23 (17.0)
Workers/farmers	26	23.15 (6.69)	21 (77.8)	5 (19.2)	19 (73.1)	7 (26.9)	23 (85.5)	3 (11.5)
Others	20	26.25 (7.62)	15 (75.0)	5 (25.0)	14 (70.0)	6 (30.0)	16 (80.0)	4 (20.0)
Total	687	23.70 (7.52)	564 (82.1)	123 (17.9)	479 (69.7)	208 (30.3)	593 (86.3)	94 (13.7)

[a] NM: normal–mild.

[b] MES: moderate–extremely severe.

Working Hours Per Day

The majority of the participants reported working 4-8 hours per day (n=427, 61.4%). This was the case in the following groups—nurses: 145/221, 65.6%; students: 39/43, 90.7%; teachers/government staff: 47/60, 78.4%; economy staff: 97/135, 71.9%; workers/farmers: 18/26, 69.3%; others: 15/20, 75.0%.

A sizeable part of the sample reported working 8-12 hours per day (n=260, 37.4%). This high workload typically affected doctors (103/158, 65.2%) and other medical staff (13/24, 54.2%).

Contact With COVID-19 at Work

In total, 6 (0.9%) participants were infected themselves (2 doctors, 3 nurses, and 1 member of the group teachers/government staff). Of all participants, 180 (26.2%) had contact with people infected by the virus at work. The most affected group were doctors (68/158, 43.0% had contact with COVID-19 at work), followed by other medical staff (10/24, 41.7%), nurses (88/221, 39.8%), teachers/government staff (8/60, 13.3%), economy staff (5/135, 3.8%), workers/farmers (1/26, 3.8%). Participants from the other professions category and students did not report contact with COVID-19 at work.

Media Preferences

When asked about the primary way participants obtained information in the past month, the majority of respondents indicated having done so through WeChat (n=606, 88.2%) (Table 3).

Table 3. Participants' answers to the multiple-choice question: what was your main way of obtaining information during the last month?

Source	Participants, n (%)
Newspaper	53 (7.71)
Television	465 (67.69)
Weibo	304 (44.25)
WeChat	606 (88.21)
Circle of friends[a]	502 (73.07)
Family/colleagues	311 (45.27)
Other	104 (15.14)

[a] Includes WeChat groups and other social media–related groups.

Coping Strategies

The three most successful ways of facing the demands of COVID-19 in daily life and work, out of 12 possible answers, were taking protective measures (mean 2.57, SD 0.67), actively acquiring more knowledge about COVID-19 (mean 2.09, SD 0.78), and engaging in recreational activities (mean 1.94, SD 0.77*)*. All coping strategies are listed in Table 4.

Three dimensions could be extracted after carrying out the PCA and were named as active coping, mental coping, and emotional coping, after analyzing the content of the items. The dimensions accounted for 47.2% of the variance (Table 4).

Table 4. Matrix of coping strategy components and three statistics after varimax rotation (the rotation is converged in five iterations; method of extraction: main component analysis).

Items	Mean (SD)	Factor loadings		
		Active	Mental	Emotional
Coping strategy components				
Taking protective measures (washing hands, wearing a mask, taking one’s own temperature, etc)	2.57 (0.67)	0.77	–0.03	0.01
Actively acquiring more knowledge about COVID-19 (symptoms, transmission pathway, etc)	2.09 (0.78)	0.75	0.17	0.02
Changing one’s thoughts and facing the situation with a positive attitude	1.90 (0.83)	0.53	0.44	–0.12
Engaging in recreational activities (WeChat, Weibo, TikTok, online shopping, online movies, exercises)	1.94 (0.77)	0.42	0.23	–0.06
Video chatting with family and friends by phone to share concerns and support	1.69 (0.80)	0.40	0.43	–0.02
Engaging in health-promoting behaviors (more rest, exercise, balanced diet, etc)	1.76 (0.82)	0.27	0.68	–0.06
Acquiring mental health knowledge and information	1.36 (0.91)	0.27	0.63	0.01
Practicing relaxation methods (meditation, yoga, Tai Chi, etc)	0.88 (0.85)	–0.07	0.83	0.09
Limiting oneself from watching too much news about COVID-19	0.53 (0.73)	–0.11	0.13	0.72
Distracting oneself from thinking about COVID-19 issues by suppression or keeping busy	0.70 (0.82)	0.03	0.10	0.75
Venting emotions by crying, screaming, smashing things, and so on	0.23 (0.48)	–0.08	–0.08	0.53
Using alcohol or drugs	0.22 (0.53)	0.07	–0.10	0.53
Statistics				
Eigenvalue	—[a]	2.89	1.71	1.07
Percentage of total variance	—	24.05	14.23	8.93
Total variance	—	—	—	47.21

[a] Not applicable.

Predictors of Increased Levels of Depression, Anxiety, and Stress

We calculated three binary logistic regression models in order to find associations of gender, lockdown area, contact with COVID-19 infection at work, and coping factors with the odds of belonging to the group for moderate–extremely severe depression, anxiety, or stress. Being female and applying emotional coping strategies increased the probability of belonging to the moderate–extremely severe depression, anxiety, or stress group. Applying active coping strategies reduced the probability of being affected by moderate–extremely severe anxiety, while mental coping strategies reduced the probability in all three moderate–extremely severe mental health groups. Age, being in a lockdown area, having contact with COVID-19 at work, and having a high workload (8-12 hr per day) did not significantly predict the odds of expressing moderate–extremely severe symptoms of depression, anxiety, or stress. The results are displayed in detail in Table 5.

Table 5. Results of a logistic regression predicting the probability of experiencing moderate–extremely severe (MES) depression, anxiety, or stress.

Variable	MES depression		MES anxiety		MES stress	
	B (SE)	OR[a] (95% CI)	B (SE)	OR (95% CI)	B (SE)	OR (95% CI)
Gender (female)	0.81 (0.27)	2.24 (1.33- 3.77)	0.47 (0.22)	1.61 (1.05-2.47)	0.78 (0.30)	2.19 (1.21-3.96)
Age	0.01 (0.01)	1.01 (0.98-1.03)	–0.01 (0.01)	0.99 (0.97-1.00)	–0.01 (0.01)	0.99 (0.96-1.01)
Hubei/Zhejiang	–0.43 (0.25)	0.65 (0.40-1.05)	–0.07 (0.21)	0.93 (0.61-1.41)	–0.22 (0.28)	0.80 (0.46-1.39)
Contact with COVID-19 infection at work	0.07 (0.26)	1.08 (0.64-1.80)	0.15 (0.49)	1.16 (0.76-1.79)	–0.03 (0.30)	0.97 (0.54-1.74)
Daily workload (8-12 hr)	0.06 (0.22)	1.06 (0.69-1.65)	0.23 (0.19)	1.26 (0.87-1.83)	0.32 (0.25)	1.38 (0.85-2.25)
Active coping	–0.14 (0.11)	0.87 (0.71-1.07)	–0.21 (0.09)	0.81 (0.68-0.97)	–0.11 (0.12)	0.89 (0.70-1.13)
Mental coping	–0.56 (0.12)	0.57 (0.45-0.72)	–0.42 (0.10)	0.67 (0.55-0.81)	–0.67 (0.14)	0.51 (0.39-0.67)
Emotional coping	0.63 (0.10)	1.89 (1.55-2.30)	0.80 (0.10)	2.16 (1.80-2.60)	0.82 (0.11)	2.27 (1.81-2.84)

[a] OR: odds ratio.

Major Stressors (Health Care Workers Only)

Out of 18 stressors, the three most demanding aspects for health care workers (n=375) were related to worries about infecting one’s family with COVID-19 (mean 1.46, SD 0.86), followed by the potential deterioration of their patients’ condition (mean 1.42, SD 0.79) and their patients’ emotional reaction (mean 1.3, SD 0.81) (Table 6).

Table 6. Doctors’ and nurses’ responses to the question: when you think about COVID-19 in your life and work, how often did you think or worry about the following things? (0=not at all, 3=very much) (n=375).

Stressor	Response, mean (SD)
Worries about infecting your family with COVID-19	1.46 (0.86)
Deterioration of patients’ condition	1.42 (0.79)
Patients’ emotional reaction	1.30 (0.81)
Emotional reaction of patients’ families	1.29 (0.79)
Uncertainties about when the epidemic will be under control	1.27 (0.78)
Coworkers displaying COVID-19–like symptoms	1.27 (0.79)
Worries about getting infected	1.24 (0.78)
Worries about being negligent and endangering patients	1.23 (0.88)
Worries about lack of proper knowledge and equipment	1.23 (0.79)
Worries about being negligent and endangering coworkers	1.18 (0.83)
Worries about nosocomial spread	1.15 (0.82)
Conflict between duty and safety	1.15 (0.81)
Being infected by colleagues	1.12 (0.81)
Protective gears being a hinderance to providing quality care	1.12 (0.80)
Being blamed by supervisors/managers	1.10 (0.80)
Displaying COVID-19–like symptoms yourself	1.09 (0.77)
Worries about the lack of manpower	1.07 (0.91)
Being without a properly equipped environment	1.05 (0.84)
Physical discomfort caused by protective gears	1.01 (0.79)
Ambiguity in the responsibilities between doctors and nurses	1.00 (0.86)
Frequent modification of infection control procedures	0.96 (0.81)
Coworkers being emotionally unstable	0.96 (0.77)
Unclear documentation and reporting procedures	0.92 (0.78)

Discussion

Predictors of Mental Health Symptoms

This survey aimed to assess the psychological burden and mental health of the Chinese population during the final stages of the lockdown, as well as to determine successful coping strategies and potentially vulnerable professional groups with specific support needs. Our results suggest that being female and, independent of gender, applying certain coping strategies increased the incidence of symptoms of depression, anxiety, and stress. Emotional coping strategies like venting emotions, consuming alcohol, or limiting oneself from information were not helpful for participants dealing with COVID-19–related psychological problems. Active strategies to cope with moderate–extremely severe anxiety, such as taking protective measures and acquiring more knowledge were more beneficial, but the most effective strategy was focusing on mental coping like relaxation techniques and gaining knowledge about mental health. Our results confirm the findings of Guo et al [16], who used another instrument to determine coping strategies but found emotion-focused strategies to worsen mental health, while cognitive and problem-focused coping strategies to be helpful. Interestingly, exposure to COVID-19 at work, living within a lockdown area, and daily workload did not play a significant role in predicting elevated symptoms.

We found no overall increased mean values in perceived stress and depression, anxiety, and stress in comparison to former (pre–COVID-19) samples (eg, compared to the perceived stress levels of patients in Hong Kong [43] or among residents in Beijing [44]) regarding mental health scores. This could be because the Chinese population had already become accustomed to the burden of COVID-19 by the end of the lockdown. The first studies in January investigating the psychological impact of the outbreak on the population reported high levels of burden [16,31,45]. Wang et al [13,14] measured DASS-21 scores twice at the beginning and at the peak of the pandemic in China and found increased anxiety in nearly 30% of participants; yet no longitudinal increases were noted. In our study, we added a third point of measurement at the final stage of the lockdown and found a considerable share of people who expressed moderate to extremely severe symptoms of depression (17.9%), anxiety (30.3%), and stress (13.7%). This result should be taken seriously as our sample revealed a higher percentage of people with increased levels of depression while stress scores were lower compared to the Wang et al [13,14] studies. We also observed a high correlation of perceived stress in the past 4 weeks and actual scores in depression. Previous research on the etiology of depression could show that stress might be one predictor for this mood disorder [46,47]. Following this line of argument, an increase of depressive symptoms by the end of the lockdown could be interpreted as part of a concerning development.

Vulnerable Groups

Some groups in our sample were more affected by symptoms of depression, anxiety, or stress than others. Students were vulnerable to moderate–extremely severe symptoms in all three categories. Outside the pandemic context, Chinese students have been reported to be affected by mental health problems due to stressful academic demands [48,49]. The consequences of the pandemic on students' lives were aggravated by infection control measures, online learning on a tight schedule, and uncertain future prospects [50]. A cross-sectional study among medical students after the lockdown revealed high levels of anxiety (38% of participants) and showed that social media played an important role in adherence to protective measures among them [29]. According to our results, the main method of getting COVID-19–related information was WeChat, directly followed by circle of friends, which in China refers to WeChat groups and other social media–related groups.

Economy staff were highly burdened by stress but did not exhibit more depression or anxiety than other groups. This result does not support that of Huang and Zhao [31], who, using different protocols, questioned enterprise employees in early February, and found depression among 20% of participants and anxiety among 34%. Increased stress levels in our sample could be explained by the fact that we put together all employees and self-employed people in the finance and IT sectors as well as salespersons in one group. Enhanced stress may be a result of the concerns to ensure livelihood for their own family during the lockdown, as shown previously [16]. Further differentiated results are needed to allow for a more specific statement.

Doctors and nurses in our sample were highly affected by anxiety; doctors had the highest workload per day. Nevertheless, perceived stress and DASS-21 stress levels were not higher than other groups, which may be due to a high professional devotion, as reported in previous research [51]. Issues about patient care like the deterioration of the patients' condition and the emotional reaction of patients and their families were perceived as stressful. In another study, nurses in Wuhan were found lacking in training for dealing with uncooperative patients [26]. Several studies have reported on the need for training Chinese medical staff on doctor-patient communication outside the context of the COVID-19 crisis [52]. The anxiety felt by doctors and nurses in our sample may be the result of the fear of bringing COVID-19 to their own families, the most intense stressor in our and other samples [23,51]. A further stressor found in other studies included spending too much time on social media while searching for COVID-19 updates. Doing so promoted symptoms of depression and anxiety in Chinese adults in general as well as in health professionals [19].

Finally, other medical staff, a small group consisting of volunteers, midwives, and pharmacists in our sample, was more affected by depression and stress than other groups and was vulnerable to anxiety as well. During the pandemic, many volunteers supported hospitals in a frontline capacity [53], and our results show that these individuals, who had the second highest workload per day after doctors, have been overlooked so far and should receive more attention since they seem to have special support needs. This group might lack the institutional psychological support that is delivered to the core personnel in a frontline hospital.

Future Implications

By 2017, there were only 33,400 licensed psychiatrists in China [54]. In recent years, the Chinese National Health Commission has established several mental health initiatives to extend mental health care to the general population. During the COVID-19 pandemic, the commission published guidelines and treatment instructions, uploaded videos of mental health education via WeChat, and established expert emergency groups for mental health services at the hospitals [55].

However, health care workers continue to be in dire need of greater access to specific mental health services [56], and further research is needed on the role of media consumption and mental health during such crises. Although the internet provided many supporting measures like telemedicine for patients with COVID-19 [57] and online mental health education and counseling [58], excessive exposure to the media seems to play a significant role in explaining mental health problems during the pandemic [18]. The Chinese population might find itself in a paradox: acquiring knowledge about the virus and mental health measures is helpful but in order to maximize coverage this knowledge is spread by governmental and other institutions via social media [55,58], which has its own harmful mental consequences [16,19]. As a result, it is not possible for individuals to avoid media consumption. More conclusive findings on the role of the media and the mental health of subgroups of the Chinese population are needed.

Limitations

Although we received 687 responses, the professional groups in this study were not of equal size; numbers were especially limited for the students and other medical staff categories, which reduces the power of statistical analysis. Targeted investigations may be needed to assess the status of underrepresented professions in a differentiated way. Further, online studies are unable to allow a valuable diagnostic assessment, and this limitation applies to all former studies. This is further aggravated due to the great variability in instruments used in different surveys, which reduces the comparability of results. Some authors used the DASS-21 previously [13,14] but with different cut off-scores (see below), while some relied on the Self-rating Depression Scale (SDS) [45], the Patient Health Questionnaire (PHQ) [56], the Generalized Anxiety Disorder 7-item (GAD-7), the Center for Epidemiology Scale for Depression (CES-D) [31], or the Post-Traumatic Stress Disorders (PTSD) Checklist [16]. However, only a face-to-face diagnostic interview by a qualified medical doctor or a clinical psychologist could allow a statement regarding a mental disorder. Therefore, our results hint at a certain development but should not be interpreted as a diagnosis of the population. Moreover, we compared our results to the studies of Wang et al [13,14], who used slightly different cut-off scores in interpreting the borders between mild and moderate depression and anxiety as suggested originally by Lovibond and Lovibond [33]. However, even without knowing their exact frequencies for single scores, our findings on increases in depression would have been even higher if Wang et al had used the original cut-off scores. Finally, we only asked for media preferences and not for the time spent on media platforms. A detailed analysis on the reasons for media consumption (entertainment, information, relaxation), the way of usage (alone, together), and mental health is necessary to reveal a more holistic picture.

Conclusion

A considerable part of the general population in China reported elevated symptoms of depression, anxiety, and stress during the final stages of the COVID-19 lockdown. Doctors, nurses, students, and other medical staff were found to be in imminent danger of developing mental health problems. Similarly, economy staff was also highly stressed. Being female was an additional risk factor for potential vulnerability toward developing mental health problems. We recommend providing additional specific information to these subgroups targeting their respective mental health profile and to personalize the successful coping strategies found in our results (ie, active and mental coping). These refer to constructive ways of behavior (eg, actively acquiring knowledge, applying protective measures) and mental health strategies (eg, relaxation techniques, psycho-education, and promoting social contact). Profession-specific mental health prevention programs should be developed and provided in formats preferred by the respective age, gender, or professional groups.

Acknowledgments

We thank all participants for taking part in our study despite the current challenges imposed by the pandemic. We also thank Wuhan CNweb Pioneer Ltd, which was responsible for the implementation of the online survey in collaboration with the QHealth Trusted Doctors Group (Penguin), Wuhan Saidian Consulting Service Company, and Tao Doctor APP. We appreciate the support of Tongji Medical College for disseminating the questionnaire link.

Conflicts of Interest

None declared.

References

1. Pneumonia of unknown cause – China. World Health Organization. 2020 Jan 5. URL: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/ [accessed 2020-08-01]
2. Timeline of WHO's response to COVID-19. World Health Organization. 2020 Jun 29. URL: https://www.who.int/news-room/detail/29-06-2020-covidtimeline [accessed 2020-08-01]

3. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
4. Desheng C. Wuhan reopens after 76-day lockdown. Chinadaily. 2020 Apr 8. URL: http://www.chinadaily.com.cn/a/202004/08/WS5e8d06fca310aeaeeed5099e.html [accessed 2020-08-01]
5. Fifield A. Hundreds of miles from Hubei, additional 30 million Chinese are in coronavirus lockdown. The Washington Post. 2020 Feb 7. URL: https://www.washingtonpost.com/world/asia_pacific/hundreds-of-miles-from-hubei-another-30-million-chinese-are-in-coronavirus-lockdown/2020/02/07/03a08282-48b9-11ea-8a1f-de1597be6cbc_story.html [accessed 2020-08-01]
6. Fiorillo A, Gorwood P. The consequences of the COVID-19 pandemic on mental health and implications for clinical practice. Eur Psychiatry 2020 Apr 01;63(1):e32 [FREE Full text] [doi: 10.1192/j.eurpsy.2020.35] [Medline: 32234102]
7. Tam CWC, Pang EPF, Lam LCW, Chiu HFK. Severe acute respiratory syndrome (SARS) in Hong Kong in 2003: stress and psychological impact among frontline healthcare workers. Psychol Med 2004 Oct 21;34(7):1197-1204. [doi: 10.1017/s0033291704002247] [Medline: 15697046]
8. Maunder R, Hunter J, Vincent L, Bennett J, Peladeau N, Leszcz M, et al. The immediate psychological and occupational impact of the 2003 SARS outbreak in a teaching hospital. CMAJ 2003 May 13;168(10):1245-1251 [FREE Full text] [Medline: 12743065]
9. Khalid I, Khalid TJ, Qabajah MR, Barnard AG, Qushmaq IA. Healthcare Workers Emotions, Perceived Stressors and Coping Strategies During a MERS-CoV Outbreak. Clin Med Res 2016 Mar 04;14(1):7-14 [FREE Full text] [doi: 10.3121/cmr.2016.1303] [Medline: 26847480]
10. Park J, Lee E, Park N, Choi YH. Mental Health of Nurses Working at a Government-designated Hospital During a MERS-CoV Outbreak: A Cross-sectional Study. Arch Psychiatr Nurs 2018 Feb;32(1):2-6 [FREE Full text] [doi: 10.1016/j.apnu.2017.09.006] [Medline: 29413067]
11. Cheung YT, Chau PH, Yip PSF. A revisit on older adults suicides and Severe Acute Respiratory Syndrome (SARS) epidemic in Hong Kong. Int J Geriatr Psychiatry 2008 Dec;23(12):1231-1238. [doi: 10.1002/gps.2056] [Medline: 18500689]
12. McAlonan GM, Lee AM, Cheung V, Cheung C, Tsang KW, Sham PC, et al. Immediate and sustained psychological impact of an emerging infectious disease outbreak on health care workers. Can J Psychiatry 2007 Apr 23;52(4):241-247. [doi: 10.1177/070674370705200406] [Medline: 17500305]
13. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. IJERPH 2020 Mar 06;17(5):1729. [doi: 10.3390/ijerph17051729]
14. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
15. Zhang W, Yang X, Zhao J, Yang F, Jia Y, Cui C, et al. Depression and Psychological-Behavioral Responses Among the General Public in China During the Early Stages of the COVID-19 Pandemic: Survey Study. J Med Internet Res 2020 Sep 04;22(9):e22227 [FREE Full text] [doi: 10.2196/22227] [Medline: 32886066]
16. Guo J, Feng X, Wang X, van IJzendoorn MH. Coping with COVID: Exposure to COVID-19 and Negative Impact on Livelihood Predict Elevated Mental Health Problems in Chinese Adults. SSRN Journal 2020 Apr 29 [FREE Full text] [doi: 10.2139/ssrn.3592667]
17. Shah K, Kamrai D, Mekala H, Mann B, Desai K, Patel RS. Focus on Mental Health During the Coronavirus (COVID-19) Pandemic: Applying Learnings from the Past Outbreaks. Cureus 2020 Mar 25;12(3):e7405 [FREE Full text] [doi: 10.7759/cureus.7405] [Medline: 32337131]
18. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020 Apr 16;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
19. Ni MY, Yang L, Leung CMC, Li N, Yao XI, Wang Y, et al. Mental Health, Risk Factors, and Social Media Use During the COVID-19 Epidemic and Cordon Sanitaire Among the Community and Health Professionals in Wuhan, China: Cross-Sectional Survey. JMIR Ment Health 2020 May 12;7(5):e19009 [FREE Full text] [doi: 10.2196/19009] [Medline: 32365044]
20. Kang L, Li Y, Hu S, Chen M, Yang C, Yang BX, et al. The mental health of medical workers in Wuhan, China dealing with the 2019 novel coronavirus. The Lancet Psychiatry 2020 Mar;7(3):e14. [doi: 10.1016/s2215-0366(20)30047-x]
21. Guo J, Liao L, Wang B, Li X, Guo L, Tong Z, et al. Psychological Effects of COVID-19 on Hospital Staff: A National Cross-Sectional Survey of China Mainland. SSRN Journal 2020 Mar 13 [FREE Full text] [doi: 10.2139/ssrn.3550050]
22. Zhang W, Wang K, Yin L, Zhao W, Xue Q, Peng M, et al. Mental Health and Psychosocial Problems of Medical Health Workers during the COVID-19 Epidemic in China. Psychother Psychosom 2020 Apr 9;89(4):242-250 [FREE Full text] [doi: 10.1159/000507639] [Medline: 32272480]
23. Hall H. The effect of the COVID-19 pandemic on healthcare workers' mental health. JAAPA 2020 Jul;33(7):45-48. [doi: 10.1097/01.JAA.0000669772.78848.8c] [Medline: 32590533]

24. Huang L, Xu F, Liu H. Emotional responses and coping strategies of nurses and nursing college students during COVID-19 outbreak. medRxiv 2020 Preprint posted online March 08, 2020. [doi: 10.1101/2020.03.05.20031898]
25. Zhuo K, Gao C, Wang X, Zhang C, Wang Z. Stress and sleep: a survey based on wearable sleep trackers among medical and nursing staff in Wuhan during the COVID-19 pandemic. Gen Psych 2020 Jun 16;33(3):e100260. [doi: 10.1136/gpsych-2020-100260]
26. Chen Q, Liang M, Li Y, Guo J, Fei D, Wang L, et al. Mental health care for medical staff in China during the COVID-19 outbreak. The Lancet Psychiatry 2020 Apr;7(4):e15-e16. [doi: 10.1016/s2215-0366(20)30078-x]
27. Yang D, Tu C, Dai X. The effect of the 2019 novel coronavirus pandemic on college students in Wuhan. Psychol Trauma 2020 Aug;12(S1):S6-S14. [doi: 10.1037/tra0000930] [Medline: 32551764]
28. Zhang Y, Zhang H, Ma X, Di Q. Mental Health Problems during the COVID-19 Pandemics and the Mitigation Effects of Exercise: A Longitudinal Study of College Students in China. Int J Environ Res Public Health 2020 May 25;17(10):3722 [FREE Full text] [doi: 10.3390/ijerph17103722] [Medline: 32466163]
29. Lin Y, Hu Z, Alias H, Wong LP. Influence of Mass and Social Media on Psychobehavioral Responses Among Medical Students During the Downward Trend of COVID-19 in Fujian, China: Cross-Sectional Study. J Med Internet Res 2020 Jul 20;22(7):e19982 [FREE Full text] [doi: 10.2196/19982] [Medline: 32584779]
30. Zhou X. Managing psychological distress in children and adolescents following the COVID-19 epidemic: A cooperative approach. Psychol Trauma 2020 Aug;12(S1):S76-S78. [doi: 10.1037/tra0000754] [Medline: 32551769]
31. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
32. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007 Oct 16;147(8):573-577 [FREE Full text] [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
33. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Sydney, NSW: Psychology Foundation of Australia; 1996.
34. Zuo K, Chang A. Zuo and Chang Chinese translation (simplified characters). Psychology Foundation of Australia. URL: http://www2.psy.unsw.edu.au/dass//Chinese/Chinese%20simplified.htm [accessed 2020-08-01]
35. Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychological Assessment 1998 Jun;10(2):176-181. [doi: 10.1037/1040-3590.10.2.176]
36. Cohen S, Kamarck T, Mermelstein R. A Global Measure of Perceived Stress. Journal of Health and Social Behavior 1983 Dec;24(4):385. [doi: 10.2307/2136404]
37. Shi C, Guo Y, Ma H, Zhang M. Psychometric validation of the 14-item perceived stress scale in Chinese medical residents. Curr Psychol 2019 May 27;38(6):1428-1434. [doi: 10.1007/s12144-019-00285-y]
38. Lee S, Juang Y, Su Y, Lee H, Lin YH, Chao C. Facing SARS: psychological impacts on SARS team nurses and psychiatric services in a Taiwan general hospital. Gen Hosp Psychiatry 2005 Sep;27(5):352-358 [FREE Full text] [doi: 10.1016/j.genhosppsych.2005.04.007] [Medline: 16168796]
39. SPSS Statistics for Windows. Armonk, NY: IBM Corp; 2016.
40. Sartorius N, Üstün TB, Lecrubier Y, Wittchen H. Depression Comorbid with Anxiety: Results from the WHO Study on Psychological Disorders in Primary Health Care. Br J Psychiatry 2018 Aug 06;168(S30):38-43. [doi: 10.1192/s0007125000298395]
41. Backhaus K, Erichson B, Plinke W, Weiber R. Multivariate Analysemethoden. Berlin, Heidelberg: Springer Gabler; 2016.
42. Fagerland MW, Hosmer DW. A goodness-of-fit test for the proportional odds regression model. Stat Med 2013 Jun 15;32(13):2235-2249. [doi: 10.1002/sim.5645] [Medline: 23037691]
43. Leung DY, Lam T, Chan SS. Three versions of Perceived Stress Scale: validation in a sample of Chinese cardiac patients who smoke. BMC Public Health 2010 Aug 25;10(1):513 [FREE Full text] [doi: 10.1186/1471-2458-10-513] [Medline: 20735860]
44. Chan RC, Xu T, Huang J, Wang Y, Zhao Q, Shum DH, et al. Extending the utility of the Depression Anxiety Stress scale by examining its psychometric properties in Chinese settings. Psychiatry Res 2012 Dec 30;200(2-3):879-883. [doi: 10.1016/j.psychres.2012.06.041] [Medline: 22921506]
45. Chen B, Li Q, Zhang H, Zhu J, Yang X, Wu Y, et al. The psychological impact of COVID-19 outbreak on medical staff and the general public. Curr Psychol 2020 Oct 07:1-9. [doi: 10.1007/s12144-020-01109-0] [Medline: 33046955]
46. Vinkers CH, Joëls M, Milaneschi Y, Kahn RS, Penninx BW, Boks MP. Stress exposure across the life span cumulatively increases depression risk and is moderated by neuroticism. Depress Anxiety 2014 Sep 17;31(9):737-745. [doi: 10.1002/da.22262] [Medline: 24753162]
47. Ménard C, Hodes G, Russo S. Pathogenesis of depression: Insights from human and rodent studies. Neuroscience 2016 May 03;321:138-162 [FREE Full text] [doi: 10.1016/j.neuroscience.2015.05.053] [Medline: 26037806]

48. Lun KW, Chan C, Ip PK, Ma SY, Tsai W, Wong C, et al. Depression and anxiety among university students in Hong Kong. Hong Kong Med J 2018 Oct 24;24(5):466-472 [FREE Full text] [doi: 10.12809/hkmj176915] [Medline: 30245480]
49. Mao Y, Zhang N, Liu J, Zhu B, He R, Wang X. A systematic review of depression and anxiety in medical students in China. BMC Med Educ 2019 Sep 02;19(1):327 [FREE Full text] [doi: 10.1186/s12909-019-1744-2] [Medline: 31477124]
50. Li HY, Cao H, Leung DYP, Mak YW. The Psychological Impacts of a COVID-19 Outbreak on College Students in China: A Longitudinal Study. Int J Environ Res Public Health 2020 Jun 02;17(11):3933 [FREE Full text] [doi: 10.3390/ijerph17113933] [Medline: 32498267]
51. Wang H, Liu Y, Hu K, Zhang M, Du M, Huang H, et al. Healthcare workers' stress when caring for COVID-19 patients: An altruistic perspective. Nurs Ethics 2020 Nov 14;27(7):1490-1500. [doi: 10.1177/0969733020934146] [Medline: 32662326]
52. Liu X, Rohrer W, Luo A, Fang Z, He T, Xie W. Doctor-patient communication skills training in mainland China: a systematic review of the literature. Patient Educ Couns 2015 Jan;98(1):3-14. [doi: 10.1016/j.pec.2014.09.012] [Medline: 25308954]
53. Kanthor R. A New Generation of Volunteers Emerges in the Chinese City at the Epicenter of the Coronavirus Crisis. The Times. 2020 Feb 24. URL: https://time.com/5789530/wuhan-volunteers-coronavirus-covid-19-china/ [accessed 2020-08-01]
54. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. The Lancet Psychiatry 2020 Apr;7(4):300-302. [doi: 10.1016/s2215-0366(20)30073-0]
55. Li W, Yang Y, Liu Z, Zhao Y, Zhang Q, Zhang L, et al. Progression of Mental Health Services during the COVID-19 Outbreak in China. Int J Biol Sci 2020;16(10):1732-1738 [FREE Full text] [doi: 10.7150/ijbs.45120] [Medline: 32226291]
56. Kang L, Ma S, Chen M, Yang J, Wang Y, Li R, et al. Impact on mental health and perceptions of psychological care among medical and nursing staff in Wuhan during the 2019 novel coronavirus disease outbreak: A cross-sectional study. Brain Behav Immun 2020 Jul;87:11-17 [FREE Full text] [doi: 10.1016/j.bbi.2020.03.028] [Medline: 32240764]
57. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine During the COVID-19 Pandemic: Experiences From Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
58. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. The Lancet Psychiatry 2020 Apr;7(4):e17-e18. [doi: 10.1016/s2215-0366(20)30077-8]

Abbreviations

CES-D: Center for Epidemiology Scale for Depression
CPSS-14: Simplified Chinese version of the 14-item Perceived Stress Scale
DASS-21: Depression, Anxiety and Stress Scale - 21 Items
GAD-7: Generalized Anxiety Disorder 7-item
KMO: Kaiser-Meyer-Olkin
MERS: Middle East respiratory syndrome
PCA: principal component analysis
PHQ: Patient Health Questionnaire
PSS-14: 14-item Perceived Stress Scale
PTSD Checklist: Post-Traumatic Stress Disorders Checklist
SARS: severe acute respiratory syndrome
SDS: Self-rating Depression Scale
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by G Eysenbach; submitted 10.09.20; peer-reviewed by C Kaylor-Hughes; comments to author 02.10.20; revised version received 13.10.20; accepted 28.10.20; published 02.12.20.

Please cite as:
Du J, Mayer G, Hummel S, Oetjen N, Gronewold N, Zafar A, Schultz JH
Mental Health Burden in Different Professions During the Final Stage of the COVID-19 Lockdown in China: Cross-sectional Survey Study
J Med Internet Res 2020;22(12):e24240
URL: https://www.jmir.org/2020/12/e24240
doi:10.2196/24240
PMID:33197231

Original Paper

The Twitter Social Mobility Index: Measuring Social Distancing Practices With Geolocated Tweets

Paiheng Xu[1], MSc; Mark Dredze[1], PhD; David A Broniatowski[2], PhD

[1]Malone Center for Engineering in Healthcare, Center for Language and Speech Processing, Department of Computer Science, Johns Hopkins University, Baltimore, MD, United States

[2]Department of Engineering Management and Systems Engineering, The George Washington University, Washington, DC, United States

Corresponding Author:
Mark Dredze, PhD
Malone Center for Engineering in Healthcare
Center for Language and Speech Processing, Department of Computer Science
Johns Hopkins University
Malone 339
3400 N Charles St
Baltimore, MD, 21218
United States
Phone: 1 4105166786
Email: mdredze@cs.jhu.edu

Abstract

Background: Social distancing is an important component of the response to the COVID-19 pandemic. Minimizing social interactions and travel reduces the rate at which the infection spreads and "flattens the curve" so that the medical system is better equipped to treat infected individuals. However, it remains unclear how the public will respond to these policies as the pandemic continues.

Objective: The aim of this study is to present the Twitter Social Mobility Index, a measure of social distancing and travel derived from Twitter data. We used public geolocated Twitter data to measure how much users travel in a given week.

Methods: We collected 469,669,925 tweets geotagged in the United States from January 1, 2019, to April 27, 2020. We analyzed the aggregated mobility variance of a total of 3,768,959 Twitter users at the city and state level from the start of the COVID-19 pandemic.

Results: We found a large reduction (61.83%) in travel in the United States after the implementation of social distancing policies. However, the variance by state was high, ranging from 38.54% to 76.80%. The eight states that had not issued statewide social distancing orders as of the start of April ranked poorly in terms of travel reduction: Arkansas (45), Iowa (37), Nebraska (35), North Dakota (22), South Carolina (38), South Dakota (46), Oklahoma (50), Utah (14), and Wyoming (53). We are presenting our findings on the internet and will continue to update our analysis during the pandemic.

Conclusions: We observed larger travel reductions in states that were early adopters of social distancing policies and smaller changes in states without such policies. The results were also consistent with those based on other mobility data to a certain extent. Therefore, geolocated tweets are an effective way to track social distancing practices using a public resource, and this tracking may be useful as part of ongoing pandemic response planning.

(J Med Internet Res 2020;22(12):e21499) doi:10.2196/21499

KEYWORDS

COVID-19; social distancing; mobility; Twitter; social media; surveillance; tracking; travel; index

Introduction

The outbreak of SARS-CoV-2, a coronavirus that causes the disease COVID-19, has caused a pandemic on a scale unseen in a generation. Without an available vaccine to reduce transmission of the virus, public health organizations and elected officials have called on the public to practice social distancing. Social distancing is a set of practices in which individuals maintain a physical distance to reduce the number of physical contacts they encounter [1,2]. These practices include maintaining a distance of at least six feet from other people and avoiding large gatherings [3]. At the time of this writing, in the United States, nearly every state had implemented statewide “stay-at-home” orders to enforce social distancing practices [4].

Social distancing is an important tool in the fight against COVID-19; however, its implementation by the general public can vary widely. Although a state governor may issue an order for the practice, individuals in different states may respond to this order in different ways. Courtemanche et al [5] showed that social distancing policies in the United States reduced the daily growth rate of COVID-19 cases. However, if we only consider the social distancing policy duration and daily confirmed cases, it is difficult to rule out potential confounders, including additional policies for wearing masks and improving hygiene as well as other social norms. Therefore, understanding actual reductions in travel and social contacts is critical to measuring the effectiveness of such policies. Using mobile phone data, Badr et al [6] found that mobility patterns were strongly correlated with decreased rates of COVID-19 case growth for the 25 most affected counties in the United States. These social distancing policies may remain in effect for an extended period of time. Thus, the public may begin to relax their practices, making additional policies necessary. Researchers showed the effectiveness of strict social distancing followed by testing and contact tracing by modeling mobility data from Cuebiq Inc in the Boston metropolitan area [7]. Additionally, epidemiologists have already modeled the impact of social distancing policies on the course of disease outbreaks [8-10]. These models may be more effective when incorporating actual measures of social distancing rather than assuming that official policies are implemented in practice.

It can be challenging to obtain data on the efficacy of social distancing practices, especially during an ongoing pandemic. In a recent Gallup poll that surveyed Americans, it was found that many adults are taking precautions to maintain distance from others [11]. However, while polling can provide insights, it cannot provide a solution. Polling is relatively expensive; thus, it is a poor choice for ongoing population surveillance practices and providing data on specific geographic locales (ie, US states and major cities) [12]. Additionally, polling around public health issues suffers from response bias, as individuals may overstate their compliance with established public health recommendations [13].

Over the past decade, analyses of social media and web data have been widely adopted to support public health objectives [14]. In this vein, several efforts have emerged over the past few months to track social distancing practices using these data sources. Google has released COVID-19 Community Mobility Reports [15] that use Google data to “chart movement trends over time by geography, across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential.” The Unacast Social Distancing Scoreboard uses data collected from 127 million monthly active users to measure the implementation of social distancing practices [16]. Researchers at the Institute for Disease Modeling have used data from Facebook’s Data for Good program to model the decline in mobility in the greater Seattle area and its effect on the spread of COVID-19 [17]. Using mobile phone data, the *New York Times* completed an analysis that showed that stay-at-home orders dramatically reduced travel; however, it was found that in states where such orders were not quickly enacted, residents continued to travel widely [18].

Identifying and isolating individuals who have potentially been exposed to a virus can blunt the spread of a pandemic. Contact tracing involves finding people who have had contact with an infected individual during the time the individual was contagious. In the current pandemic, several efforts have been made to develop digital contact tracing tools. Google and Apple announced a joint effort to build a Bluetooth-based contact tracing platform, which enhances the interoperability between Android and IOS devices using apps from public health authorities [19]. Singapore [20] and Australia [21] released similar apps that use Bluetooth to exchange “digital handshakes” to establish contacts. Many countries have developed their own contact tracing responses [22]. Li and Guo [23] presented a review of the development of contact-tracing apps for COVID-19. These efforts provide new and important opportunities to study social distancing and contact tracing in real time.

We present the Twitter Social Mobility Index, a measure of social distancing and travel patterns derived from public Twitter data. We used public geolocated Twitter data to measure how much a user travels in a given week. We computed a metric based on the standard deviation of a user’s geolocated tweets each week, and we aggregated these data over an entire population to produce a metric for the United States as a whole, for individual states, and for some US cities. We found that in the United States as a whole, there was a dramatic drop in travel in the later weeks of the study period, with travel between March 16 and April 27, 2020, showing the lowest amount since January 1, 2019, the start of our data set. Additionally, we found that travel reductions were not uniform across the United States but varied from state to state. However, there was no clear correlation between social mobility and confirmed COVID-19 cases at the state level. A key advantage of our approach is that unlike the other travel and social distancing analyses referenced above, we rely on entirely public data, which enables others to replicate our findings and explore different aspects of these data. Additionally, because Twitter contains user-generated content in addition to location information, future analyses can correlate users’ attitudes, beliefs, and behaviors with changes in social mobility.

One concern regarding the mining of social media data is user privacy. Unlike the data used by the companies described above, all the data we used is publicly available. Users choose to post their location data to Twitter publicly; therefore, these data are accessible to all users. However, while the location data are public, the potential remains for violating user privacy and producing unintended consequences for users, such as highlighting users who are failing to social distance. To ensure privacy in our index, we aggregated all mobility metrics to produce population-level analyses. None of our work considers the identity of individual users, and we removed identifiable user information from the distributed data aggregations. Furthermore, we caution others who pursue work similar to ours to consider privacy ramifications for users when collecting new data and conducting similar analyses.

There is widespread recognition that real-time tweets from millions of users can yield insights into a variety of population-level trends. Our study follows a tradition of using this insight to develop population-level indices and measures from Twitter data. Previous work includes tracking population-level sentiment as an economic indicator that can track stock price [24], political indices that reflect the popular opinion on major socioeconomic issues [25] or opinions about political candidates [26,27], and measures of pop culture such as reception of entertainment programs [28]. The Twitter Social Mobility Index is a measure of this kind, aggregating Twitter data from millions of people to produce real-time measurements of social distancing.

There is a long line of work on geolocation prediction for Twitter, which requires inferring a location for a specific tweet or user [29-32]. This includes work on patterns and trends in geotagged Twitter data [33]. Although most of these works focus on inferences of users' current locations and thus are not suitable for tracking user movements, there may be opportunities to combine these methods with our approach.

Many studies have analyzed Twitter geolocation data to study population movements. Hawelka et al [34] demonstrated a method for computing global travel patterns from Twitter, and Dredze et al [35] adapted this method to support efforts in combating the Zika virus epidemic. Several studies have used human mobility patterns from Twitter data [36-39]. These studies include analyses of urban mobility patterns [40-42]. Finally, some of these analyses considered mobility patterns around mass events [43].

Our findings are presented on a website [44], and we will continue to update our analysis during the COVID-19 pandemic.

Methods

Data Source

Twitter offers several ways in which a user can indicate their location. If a user is tweeting from a GPS-enabled device, they can attach their exact coordinates to that tweet. Twitter can then display the specific place that corresponds to these coordinates to the user and also provide it in their application programming interface (API). Alternatively, a user can explicitly select a location, which can be a point of interest (eg, a coffee shop), neighborhood, city, state, or country. If the tweet is public, this geolocation information is supplied with the tweet.

We used the Twitter streaming API [45] to download tweets based on location. We used a bounding box that covered the entire United States, including US territories. We used data from this collection starting on January 1, 2019, and ending on April 27, 2020. In total, the data set included 3,768,959 Twitter users and 469,669,925 tweets posted in the United States.

Location Data

We processed the two types of geolocation information described in the previous section.

Coordinates

We processed the exact coordinates (latitude and longitude) provided by the user (the "coordinates" field in the Twitter JavaScript Object Notation [JSON] object). Approximately 8% of our data included coordinates.

Place

The "place" field in the Twitter JSON object indicates a known location in which the tweet was authored. A place can be a point of interest (eg, a specific hotel), a neighborhood (eg, downtown Jacksonville), a city (eg, Kokomo, IN), a state (eg, Arizona), or a country (eg, the United States). The place object contains a unique ID, a bounding box, a country, and a name. More information about the location is available from the Twitter Geo API. A place is provided with a tweet in either of two conditions. First, Twitter can identify the coordinates provided by the user as occurring in a known place. Second, the user can manually select a place when authoring the tweet.

Because coordinates give a more precise location, we used them instead of place when available. If only a place was available, we assumed that the user was in the center of the place, as given by the place's bounding box.

For points of interest and neighborhoods, Twitter only provides the country in the associated metadata. Although in some cases, the city can be parsed from the name and the state inferred, we opted to exclude these places from our analysis for states. The full location details can be obtained from querying the Twitter API; however, due to the magnitude of the data in our analysis, this task would have been too time-consuming. This limitation excluded approximately 1.8% of our data.

We performed analyses for the 50 most populous US cities. For these analyses, we included points of interest that c the city name in their names, such as "New York City Center." Specifically for New York City, we included places that corresponded to each of the five New York City boroughs (Brooklyn, Manhattan, Queens, Staten Island, and the Bronx).

In summary, for each geolocated tweet, we obtained an associated latitude and longitude.

Computing Mobility

We defined the Twitter Social Mobility Index as follows. For each user, we collected all locations (coordinates) in a 1-week period, where a week starts on Monday and ends the following Sunday. We denoted the coordinate sequence as , where C_j is the coordinate at time j in week i and n is the number of coordinates in that week. We computed the centroid of all of the coordinates and considered this the "home" location for the user. We then measured the distance between each location and the centroid for that week. To determine distance, we measured the geodesic distance in kilometers between two adjacent records, C_j and C_{j+1}, using geopy [46], resulting in a distance sequence of . After collecting the distances, we measured the standard deviations of these distances. Formally, we defined Twitter Social Mobility Index M for each user as where σ(·) is the standard deviation operator and *N* is the number of weeks considered for the measure. We measured mobility in kilometers.

In summary, this measure reflects the area and regularity of travel for a user rather than the raw distance traveled. Therefore, a user who takes a long trip with a small number of check-ins would have a larger social mobility measure than a user with many check-ins who traveled in a small area. Because the measure is sensitive to the number of check-ins, it reflects when people have fewer check-ins during the pandemic.

We aggregated the results by week by taking the mean measure of all users in a given geographic area. We also present results for a 7-day moving average aggregation as a measure of daily movement. We recorded the variance of these measures to study the travel variance in the population, which indicates if travel is reduced overall but not for some users.

We produced aggregate scores by geographic area for the United States as a whole, for each US state and territory, and for the 50 most populous cities in the United States. We determined the geographic area of a user based on their centroid location for all times in our collection.

We computed the social mobility index for each day and week between January 1, 2019, and April 27, 2020. We selected the date of March 16, 2020, as the start of social distancing on the national level, although individual states implemented practices at different times. Therefore, we divided the data into two time periods: before social distancing (January 1, 2019, to March 15, 2020) and after social distancing (March 16, 2020, to April 27, 2020).

We then computed the group level reduction in social mobility by considering the average values as follows:

We also computed the reduction for each user and then tracked the median value, number of users active in both periods, and proportion of active users who completely reduced their mobility. We conducted a similar analysis for seasonal effects by comparing mobility after social distancing with mobility during the same period in 2019.

To address sparse data issues in our data set, we excluded users with fewer than 3 geolocated tweets overall and excluded the weekly record for a user if they had fewer than 3 geolocated tweets in that week. Additionally, due to data loss in our data collection process, we removed two weeks that contained far less data than the other time periods by taking a 99.75% confidence limit on the number of users and records.

Results

Social Mobility Index

Table 1 shows the Twitter Social Mobility Index measured in kilometers for every state and territory in the United States and the United States as a whole. City results are shown in Table 2. We also included the rank of location by the group level reduction.

Table 1. Reductions of mobility for all US states and territories and for the United States. Ranks are based on group level reduction.

Location	Mobility (kilometers)		Group level reduction (%)	User-level reduction (%)		Rank
	Before distancing	After distancing		Median reduction	Median seasonal reduction	
AK	109.76	25.47	76.80	99.84	63.73	1
AL	48.04	22.57	53.03	84.47	72.94	47
AR	50.54	23.15	54.19	91.87	76.81	45
AZ	62.85	23.47	62.66	93.69	85.55	26
CA	78.58	29.60	62.33	96.65	91.35	29
CO	72.23	24.47	66.12	98.2	93.37	12
CT	45.51	14.89	67.28	96.29	89.25	8
DC	77.67	19.74	74.58	100.00	97.75	2
DE	43.63	13.61	68.81	93.44	85.08	7
FL	76.99	32.24	58.13	92.38	82.92	42
GA	65.64	27.11	58.70	85.26	78.00	39
HI	147.61	70.75	52.07	97.69	89.21	51
IA	50.42	20.59	59.17	95.91	89.82	37
ID	70.77	33.36	52.86	94.12	78.19	49
IL	55.59	19.38	65.15	98.71	93.01	16
IN	45.86	17.15	62.60	97.19	89.61	27
KS	65.50	23.19	64.60	97.03	81.57	19
KY	44.67	15.31	65.74	93.93	83.42	13
LA	45.98	19.39	57.83	86.13	77.76	43
MA	58.69	17.64	69.95	98.83	93.93	5
MD	46.10	15.19	67.04	94.80	88.67	9
ME	59.68	22.45	62.38	93.77	78.53	28
MI	56.24	20.96	62.72	96.84	90.42	25
MN	64.01	21.68	66.13	98.36	91.34	11
MO	52.27	20.08	61.59	95.89	88.65	31
MS	50.24	24.36	51.51	79.09	69.11	52
MT	69.93	32.96	52.86	90.17	65.58	48
NC	52.11	19.73	62.14	94.27	85.26	30
ND	65.77	23.65	64.04	99.71	97.21	22
NE	55.11	21.88	60.29	99.95	91.40	35
NH	55.09	19.48	64.64	96.26	85.35	18
NJ	49.27	14.62	70.33	97.28	93.41	4
NM	58.20	24.23	58.37	95.66	73.14	41
NV	80.25	33.19	58.64	93.42	85.00	40
NY	71.17	24.57	65.48	98.94	94.20	15
OH	44.88	15.73	64.95	94.81	88.68	17
OK	52.34	24.69	52.83	88.38	76.99	50
OR	71.12	25.97	63.49	97.51	92.68	24
PA	54.40	19.45	64.24	97.59	89.85	20
PR	44.96	14.94	66.77	97.26	90.38	10
RI	46.80	14.50	69.01	96.74	90.55	6

Location	Mobility (kilometers)		Group level reduction (%)	User-level reduction (%)		Rank
	Before distancing	After distancing		Median reduction	Median seasonal reduction	
SC	48.28	19.85	58.88	86.03	77.92	38
SD	68.41	31.52	53.92	95.91	86.66	46
TN	56.77	21.83	61.55	94.89	85.89	32
TX	73.24	28.60	60.95	93.81	84.18	34
UT	68.43	23.62	65.49	93.56	91.50	14
VA	57.37	22.33	61.07	95.62	87.51	33
VI	132.16	47.57	64.00	98.66	87.72	23
VT	56.84	20.33	64.23	96.35	86.70	21
WA	75.34	21.31	71.71	98.43	95.72	3
WI	56.32	22.68	59.74	96.88	91.75	36
WV	46.59	20.02	57.02	88.95	82.40	44
WY	71.64	44.03	38.54	84.95	50.90	53
United States	65.59	25.04	61.83	95.86	88.36	N/A[a]

[a] N/A: not applicable.

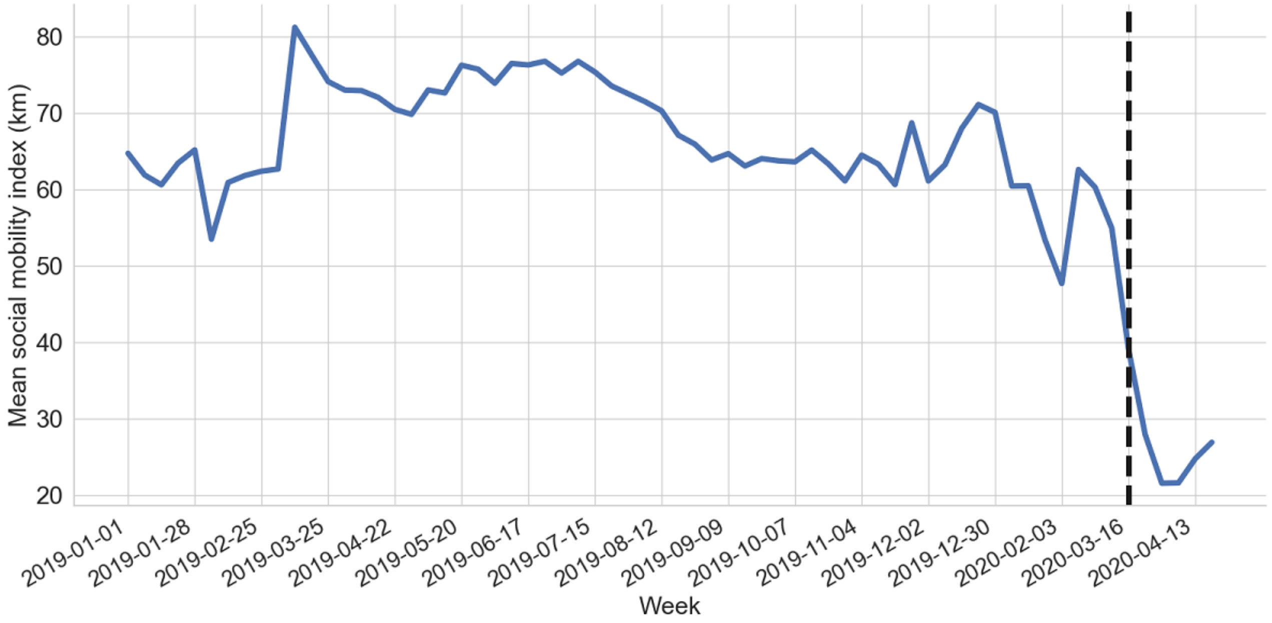
Table 2. Reduction of mobility for top 50 United States cities by population. Ranks are based on group level reduction.

Location	Mobility (kilometers)		Group level reduction (%)	User level reduction (%)		Rank
	Before distancing	After distancing		Median reduction	Median seasonal reduction	
New York City	86.37	29.91	65.38	99.70	96.69	27
Los Angeles	103.16	40.86	60.39	98.69	93.87	40
Chicago	64.09	19.87	69.00	99.96	94.58	14
Houston	53.70	21.50	59.96	97.04	88.00	41
Phoenix	60.07	19.12	68.17	96.32	91.08	18
Philadelphia	54.80	17.70	67.71	99.16	93.70	19
San Antonio	45.43	15.93	64.93	99.00	91.33	28
San Diego	79.21	28.19	64.41	98.67	92.77	30
Dallas	63.92	21.85	65.81	95.48	89.32	25
San Jose	60.63	14.82	75.55	99.88	97.34	2
Austin	72.50	22.84	68.50	99.66	94.66	17
Jacksonville	47.06	26.87	42.90	96.60	92.92	50
Fort Worth	51.67	19.68	61.92	95.33	85.72	37
Columbus	44.67	14.73	67.02	96.91	93.15	22
San Francisco	113.77	31.99	71.89	99.93	98.94	8
Charlotte	58.13	20.90	64.04	96.26	89.83	31
Indianapolis	46.50	14.53	68.76	99.26	91.85	15
Seattle	98.92	21.64	78.12	99.98	99.06	1
Denver	81.11	23.08	71.55	99.05	96.30	9
Washington	80.26	22.12	72.43	99.93	97.27	7
Boston	77.58	27.47	64.59	99.42	96.40	29
El Paso	51.10	21.50	57.92	100.00	95.97	44
Detroit	53.94	22.38	58.50	94.89	83.68	43
Nashville	72.83	23.94	67.13	98.45	94.88	21
Portland	78.91	24.81	68.56	99.45	96.81	16
Memphis	48.64	18.41	62.15	98.65	86.75	35
Oklahoma City	46.07	16.78	63.57	91.34	75.19	33
Las Vegas	80.21	35.69	55.50	94.87	83.90	47
Louisville	45.52	12.97	71.51	94.31	77.68	10
Baltimore	45.61	11.66	74.43	96.10	89.37	4
Milwaukee	52.01	22.78	56.19	97.01	91.86	46
Albuquerque	51.04	16.88	66.93	98.95	75.81	23
Tucson	53.58	23.10	56.89	95.73	84.48	45
Fresno	37.39	10.84	71.02	96.06	89.20	11
Mesa	48.77	21.72	55.47	92.40	71.33	48
Sacramento	62.14	25.45	59.05	94.82	94.47	42
Atlanta	87.90	33.39	62.02	93.50	86.36	36
Kansas City	62.93	17.23	72.61	98.30	96.54	6
Colorado Springs	64.82	23.55	63.67	99.47	95.66	32
Miami	114.33	55.77	51.22	97.55	88.56	49
Raleigh	51.62	15.24	70.47	97.79	89.51	12

Location	Mobility (kilometers)		Group level reduction (%)	User level reduction (%)		Rank
	Before distancing	After distancing		Median reduction	Median seasonal reduction	
Omaha	49.99	15.38	69.24	100.00	93.72	13
Long Beach	54.97	20.51	62.70	93.33	89.75	34
Virginia Beach	48.91	18.92	61.33	96.35	88.38	39
Oakland	87.36	22.26	74.52	98.41	96.26	3
Minneapolis	69.67	18.72	73.14	99.14	94.21	5
Tulsa	48.54	18.51	61.85	99.89	93.20	38
Arlington	56.42	18.27	67.62	97.58	93.25	20
Tampa	70.50	23.55	66.59	94.48	83.23	24
New Orleans	55.96	19.18	65.73	97.00	88.75	26

We observed that the overall drop in mobility across the United States was large (61.83%). Figure 1 shows the weekly social mobility index for the United States for the entire time period of our data set. The figure reflects a massive drop in mobility starting in March, and the four most recent weeks showed the lowest mobility on record in our data set. Second, every US state and territory saw a drop in mobility, ranging from 38.54% to 76.80% of travel compared to the numbers before March 16, 2020. However, the variance by state was high. States that were early adopters of social distancing practices ranked highly on the reduction in travel, such as Washington (3) and Maryland (9). In contrast, the eight states that had not implemented statewide orders as of the start of April [4] ranked poorly, namely Arkansas (45), Iowa (37), Nebraska (35), North Dakota (22), South Carolina (38), South Dakota (46), Oklahoma (50), Utah (14), and Wyoming (53). We observed similar trends in the city analysis; however, the median users in cities had a larger mobility reduction than the users in states.

Figure 1. Mean social mobility index (kilometers) in United States from January 1, 2019, to April 27, 2020. Weeks with missing data are excluded from the figure.

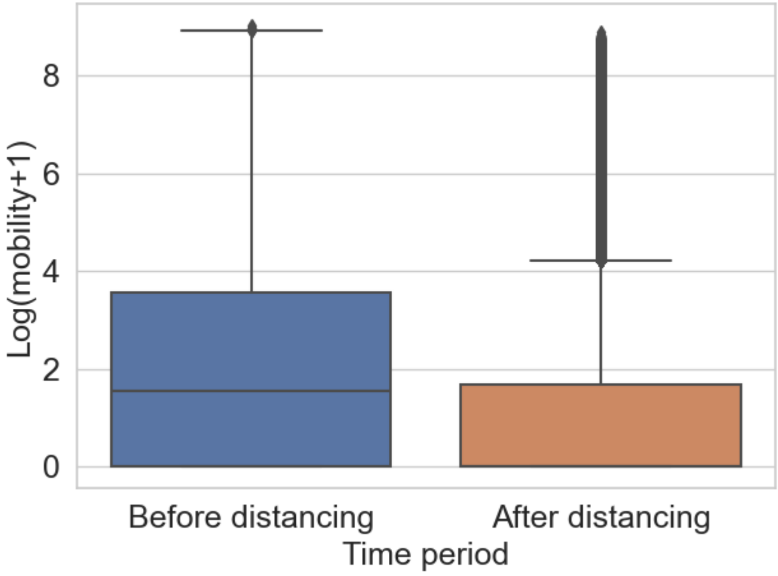


In addition to the group-level mobility travel reduction, we examined the distribution of user-level travel reduction. For this analysis, we only considered the subgroup of users who had at least two check-ins in both periods. The median values for the reduction distribution were close to 100% for most states. The median values for seasonal reduction were all smaller but still suggested that people substantially reduced their mobility during the pandemic. Moreover, in the United States, 40% of the 818,213 active users completely reduced their mobility (ie, the mobility reduction was 100%). In contrast, during the same period in 2019, a 31% reduction was seen among 286,217 active users.

The White House announced "Slow the Spread" guidelines for persons to take action to reduce the spread of COVID-19 on March 16, 2020 [47]. Of the states, 49.06% (26/53) had their largest mobility drop in the week of March 16-22, 2020, and 22.64% (12/53) had their largest drop in the following week. We computed a moving average of daily mobility data and used an offline change point detection method [48] on this trend. In 2020, 62.26% of the change points occurred after the national announcement date but before the dates on which individual state policies were enacted. This suggests that the national announcement had a larger effect compared to state policies, which is a similar finding to that of a mobile phone–based mobility analysis of four large cities [49]. We also observed that among the 40 states that announced stay-at-home policies, 92.5% (37) of the states had a more stationary daily mobility time series before the policy announcement date compared to the mobility time series over the entire time period, suggesting a rapid mobility change during the pandemic.

Finally, Figure 2 shows a box plot of the mobility variance across all users in a given time period. The distribution is long-tailed with numerous zeros; therefore, we took the log of 1 plus each mobility index. Although mobility was reduced in general, some users still showed a lot of movement, which suggests that social distancing is not being uniformly practiced. These results clearly demonstrate that our metric can track drops in travel, suggesting that it can be used as part of ongoing pandemic response planning.

Figure 2. Box plots showing the user distributions of the mean social mobility index (kilometers) before and after social distancing measure were enacted in the United States.



Correlations

To investigate the factors that explain our Twitter Social Mobility Index and how well the index tracks COVID-19 cases compared to other relevant factors, we performed a correlation analysis on our data. We computed the daily infection rate by dividing the number of new confirmed COVID-19 cases in each US state [50] by the population of the state. We compared the daily infection rate with the social mobility index and the trends in the state characteristics category from [51]. We first ran a correlation analysis for the following trends: state size in square miles, population density per square mile, unemployment rate (2018), percentage of the population living under the federal poverty line (2018), number of homeless individuals (2019), percentage of the population at risk for serious illness due to COVID-19, and number of all-cause deaths (2016). We selected these measures to track the size of the state, economic activity, and composition of the population, which were studied in a similar correlation analysis of other countries [52]. These measures may change how far people typically travel in a given state.

In Figure 3 and Figure 4, we show the characteristics that have high correlation with either the number of confirmed cases or the mobility index. These characteristics were the size of the state in square miles, the number of homeless individuals (2019), the unemployment rate (2018), and the percentage of the population at risk for serious illness due to COVID-19.

For each day, we computed the correlations between the daily infection rate and the above data by state.

Figure 3. Pearson correlations between daily COVID-19 infection rates and various factors at the state level.

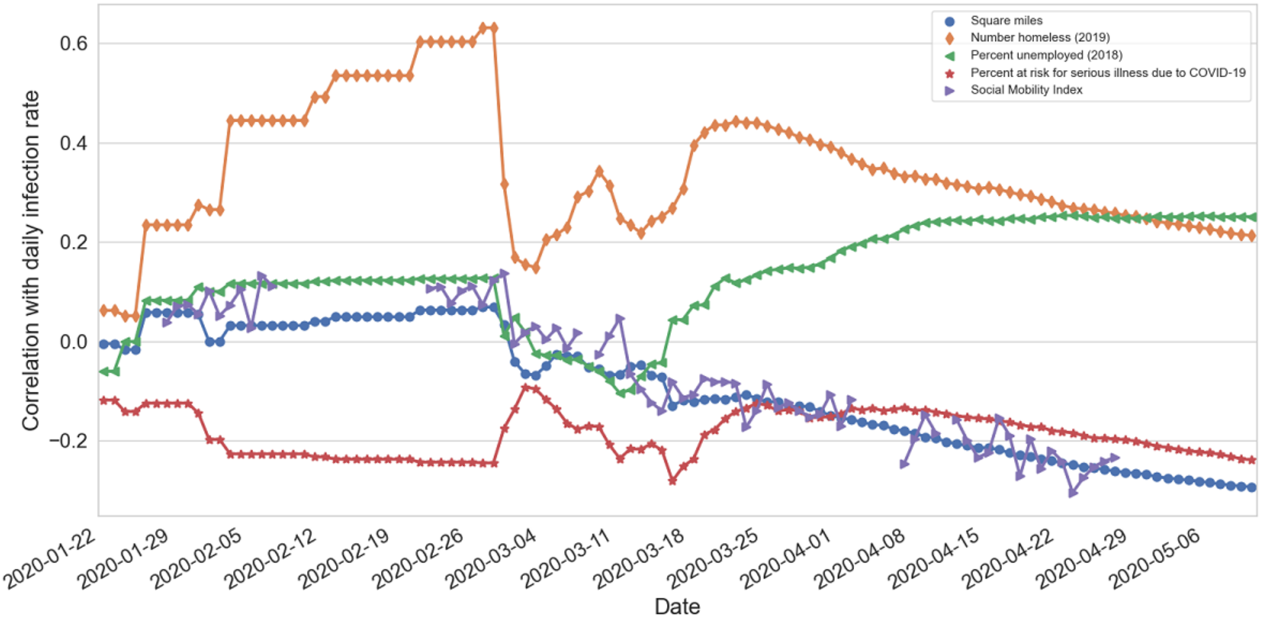


Figure 3 shows the correlations by day. We adopted the infection rate because the raw number of confirmed cases is not as informative, as the population has the highest correlation. However, the most significant factors in the early stage were still population-related factors (eg, the number of homeless people). We did not see significant correlations with other factors, including the social mobility index. Starting from mid-March, we observed trends of increasing correlation with the unemployment rate, size of the state, and social mobility index; however, these correlations were not significant (absolute correlation values <.5). A fluctuation occurred in the middle of the period, when states started to report confirmed cases of COVID-19.

We conducted a similar correlation analysis between each data source and the social mobility index, as shown in Figure 4. As expected, geographical state size showed the highest positive correlation. We also observed that the number of people at risk for serious illness due to COVID-19 had a negative correlation at the early stage of the pandemic.

Figure 4. Pearson correlations between the social mobility index and various factors at the state level.

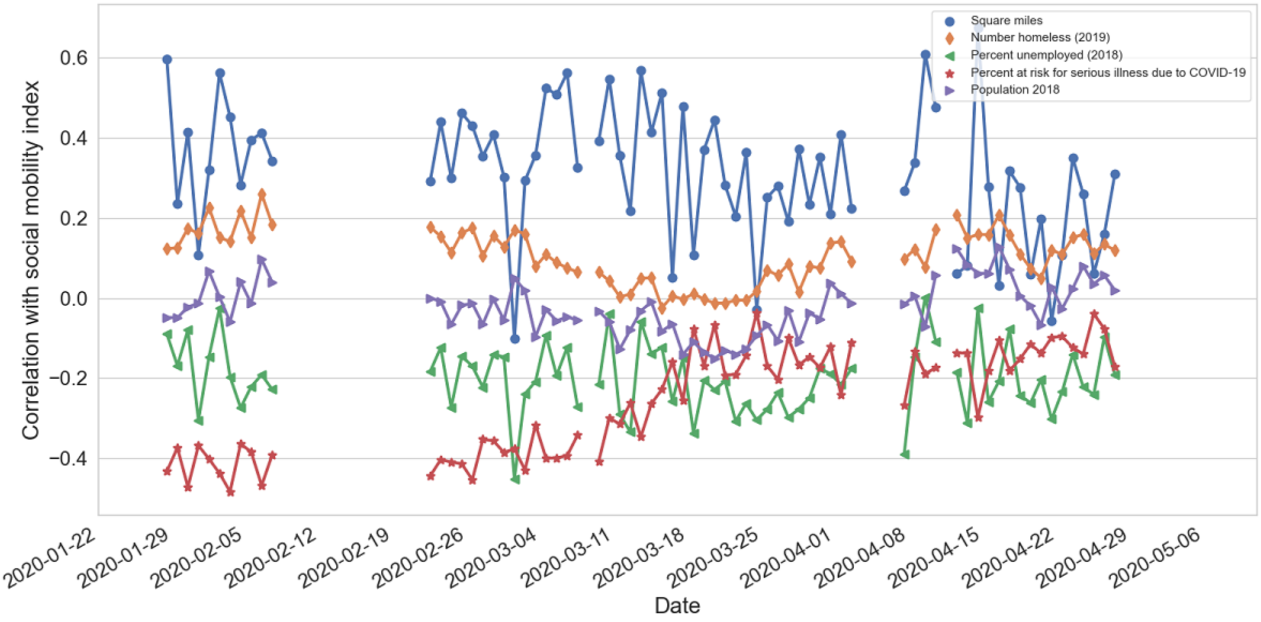


Table 3 demonstrates the effects of various restriction policies on confirmed cases by running a similar correlation analysis on the cumulative confirmed cases for each state on May 10, 2020. The policy types follow the data from [51]. We used the time difference (in days) between May 10, 2020, and the policy release date as the input for the analysis, and we assigned a negative value (–1000) to states that had not announced a policy. The factor with the highest correlation with the social mobility index is the declaration of a state of emergency, which is the broadest type of policy.

Table 3. Pearson correlations between the cumulative number of confirmed COVID-19 cases on May 10, 2020, and the dates on which policies were released in each state.

Policy	Correlation	*P* value
State of emergency	0.2587	.07
Date banned visitors to nursing homes	0.151	.29
Stay-at-home or shelter-in-place order	0.1507	.29
Evictions frozen	0.1411	.32
Nonessential businesses closed	0.1359	.34
Gyms closed	0.0765	.59
Movie theaters closed	0.0737	.61
Day cares closed	0.0563	.70
Restaurants closed except takeout	0.0341	.81
Kindergarten to 12th grade schools closed	–0.0821	.57

Discussion

We present the Twitter Social Mobility Index, a measure of social mobility based on public geolocated tweets. Our analysis shows that there was a large drop in mobility overall in the United States. However, the drop was inconsistent and varied significantly by state. It appears that states that were early adopters of social distancing practices experienced more significant drops than states that had not yet implemented these practices.

Several limitations of using geo-tagged tweets as the subject of our study must be kept in mind. First, users on Twitter and other social media platforms are not representative of the general population. Their demographics, such as age, race, ethnicity, education level, income, and political affiliation, do not perfectly mirror the larger population. In the United States, Twitter users are younger, more educated, have higher incomes, and are more likely to identify as Democrats than the general public [53,54]. Therefore, while our sample of users is large, it is highly biased.

Second, not all users are equally likely to use geotagging features on Twitter, and they may use the features in different ways. For example, in a previous study [32], demographic differences were found in the groups of people who used the two different types of geolocation information (ie, coordinates and place). GPS-tagged tweets are posted more often by young people and by women compared to tweets with self-reported locations.

Third, while we obtained access to millions of geotagged tweets, this is still a relatively small proportion of the total number of nongeotagged tweets on the platform, and it is also small compared to private measures of social mobility computed by companies such as Google and Apple.

Fourth, a small proportion of geotagged tweets report fake geolocation information. However, we believe that this is a negligible problem, as previous work found the rate of fake geolocation to be around 0.22% on social media in general [55] and even lower on Twitter. In our preliminary analysis, we considered mobility data based on GPS from mobile devices alone while excluding place information, as this method has greater precision. However, our results with these limited data were similar to our results with the full data set, except that they were less stable. Therefore, we decided to include all location data.

Despite these limitations, our results produced metrics that align with expected trends given national social distancing guidelines and related statewide policies. This suggests that there is sufficient information in our data to overcome these limitations. Additionally, the public nature of Twitter data has advantages over proprietary and private data sources. More work is needed to compare our mobility trends with those of other data sources.

Our work on this data is ongoing, and there are several directions that warrant further study. First, as states begin to reopen and some states maintain restrictions, tracking changes in population behaviors will be helpful in making policy decisions. Second, we focused on the United States; however, Twitter data provides sufficient coverage to replicate our analysis for many countries. Third, tweet content exists for each user in the data set; this content can reflect the user's attitudes, beliefs, and behaviors. Studying these factors together with users' mobility reduction could yield further insights. Our findings are presented on a website [44], and we will continue to update our analysis during the pandemic.

Conflicts of Interest

MD holds equity in Sickweather Inc and has received consulting fees from Bloomberg LP and Good Analytics Inc. These organizations did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the article. All other authors have no conflicts to declare.

References

1. Maharaj S, Kleczkowski A. Controlling epidemic spread by social distancing: do it well or not at all. BMC Public Health 2012 Aug 20;12:679 [FREE Full text] [doi: 10.1186/1471-2458-12-679] [Medline: 22905965]
2. Kelso JK, Milne GJ, Kelly H. Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza. BMC Public Health 2009 Apr 29;9(1):117 [FREE Full text] [doi: 10.1186/1471-2458-9-117] [Medline: 19400970]
3. Glass R, Glass L, Beyeler W, Min H. Targeted social distancing design for pandemic influenza. Emerg Infect Dis 2006 Nov;12(11):1671-1681 [FREE Full text] [doi: 10.3201/eid1211.060255] [Medline: 17283616]
4. Zeleny J. Why these 8 Republican governors are holding out on statewide stay-at-home orders. CNN. 2020 Apr 04. URL: https://www.cnn.com/2020/04/04/politics/republican-governors-stay-at-home-orders-coronavirus/index.html [accessed 2020-10-27]
5. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate. Health Affairs 2020 Jul 01;39(7):1237-1246. [doi: 10.1377/hlthaff.2020.00608]
6. Badr H, Du H, Marshall M, Dong E, Squire M, Gardner L. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. Lancet Infect Dis 2020 Nov;20(11):1247-1254. [doi: 10.1016/s1473-3099(20)30553-3]
7. Aleta A, Martín-Corral D, Pastore Y Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nat Hum Behav 2020 Sep 05;4(9):964-971. [doi: 10.1038/s41562-020-0931-9] [Medline: 32759985]
8. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health 2020 May;5(5):e261-e270 [FREE Full text] [doi: 10.1016/S2468-2667(20)30073-6] [Medline: 32220655]
9. Fenichel EP, Castillo-Chavez C, Ceddia MG, Chowell G, Parra PAG, Hickling GJ, et al. Adaptive human behavior in epidemiological models. Proc Natl Acad Sci USA 2011 Apr 12;108(15):6306-6311. [doi: 10.1073/pnas.1011250108] [Medline: 21444809]
10. Caley P, Philp DJ, McCracken K. Quantifying social distancing arising from pandemic influenza. J R Soc Interface 2008 Jun 06;5(23):631-639 [FREE Full text] [doi: 10.1098/rsif.2007.1197] [Medline: 17916550]
11. Saad L. Americans Step Up Their Social Distancing Even Further. Gallup. 2020 Mar 24. URL: https://news.gallup.com/opinion/gallup/298310/americans-step-social-distancing-even-further.aspx [accessed 2020-10-28]
12. Dredze M, Broniatowski DA, Smith MC, Hilyard KM. Understanding Vaccine Refusal: Why We Need Social Media Now. Am J Prev Med 2016 Apr;50(4):550-552 [FREE Full text] [doi: 10.1016/j.amepre.2015.10.002] [Medline: 26655067]
13. Adams A, Soumerai S, Lomas J, Ross-Degnan D. Evidence of self-report bias in assessing adherence to guidelines. Int J Qual Health Care 1999 Jun 01;11(3):187-192. [doi: 10.1093/intqhc/11.3.187] [Medline: 10435838]
14. Paul MJ, Dredze M. Social Monitoring for Public Health. Synthesis Lectures on Information Concepts, Retrieval, and Services 2017 Aug 31;9(5):1-183. [doi: 10.2200/s00791ed1v01y201707icr060]
15. COVID-19 Community Mobility Report. Google. URL: https://www.google.com/covid19/mobility/ [accessed 2020-10-28]
16. Social Distancing Scoreboard. Unacast. URL: https://www.unacast.com/covid19/social-distancing-scoreboard [accessed 2020-10-28]
17. Burstein R, Hu H, Thakkar N, Schroeder A, Famulare M, Klein D. Understanding the Impact of COVID-19 Policy Change in the Greater Seattle Area using Mobility Data. Institute for Disease Modeling. 2020 Mar 29. URL: https://covid.idmod.org/data/Understanding_impact_of_COVID_policy_change_Seattle.pdf [accessed 2020-10-28]
18. Glanz J, Carey B, Holder J, Watkins D, Valentino-DeVries J, Rojas R, et al. Where America Didn't Stay Home Even as the Virus Spread. New York Times. URL: https://www.nytimes.com/interactive/2020/04/02/us/coronavirus-social-distancing.html [accessed 2020-10-28]
19. Apple and Google partner on COVID-19 contact tracing technology. Apple Newsroom. URL: https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/ [accessed 2020-10-28]
20. Koh D. Singapore government launches new app for contact tracing to combat spread of COVID-19. Mobi Health News. 2020 Mar 20. URL: https://www.mobihealthnews.com/news/asia-pacific/singapore-government-launches-new-app-contact-tracing-combat-spread-covid-19 [accessed 2020-10-28]
21. Coronavirus: Australians download COVIDSafe contact tracing app. BBC. 2020 Apr 06. URL: https://www.bbc.com/news/world-australia-52433340 [accessed 2020-10-28]
22. Hale T, Webster S, Petherick A, Phillips T, Kira B. Which countries do COVID-19 contact tracing? Our World in Data. URL: https://ourworldindata.org/grapher/covid-contact-tracing [accessed 2020-10-28]
23. Li J, Guo X. Global Deployment Mappings and Challenges of Contact-tracing Apps for COVID-19. SSRN Journal Preprint posted online on May 24, 2020. [doi: 10.2139/ssrn.3609516]
24. Si J, Mukherjee A, Liu B, Li Q, Li H, Deng X. Exploiting topic based twitter sentiment for stock prediction. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. 2013 Aug Presented at: 51st Annual Meeting of the Association for Computational Linguistics; August 4-9, 2013; Sofia, Bulgaria.

25. Bollen J, Mao H, Pepe A. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2826/3237
26. Using Twitter To Graph Public Opinion of Presidential Candidates. Twitter Sentiment Index. URL: http://www.thetsindex.com/ [accessed 2020-10-28]
27. Tumasjan A, Sprenger T, Sandner P, Welpe I. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. Fourth international AAAI conference on weblogs and social media; 2010 May 16; Atlanta, Georgia, USA; 2010 Presented at: Fourth International AAAI Conference on Weblogs and Social Media; Washington, DC; May 23–26, 2010 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1441/1852
28. Hennig-Thurau T, Wiertz C, Feldhaus F. Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. J Acad Mark Sci 2014 Jun 1;43(3):375-394. [doi: 10.1007/s11747-014-0388-3]
29. Dredze M, Paul M, Bergsma S, Tran H. Carmen: A Twitter Geolocation System with Applications to Public Health. In: Expanding the Boundaries of Health Informatics Using Artificial Intelligence: Papers from the AAAI 2013 Workshop. 2013 Presented at: Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence; July 14-15, 2013; Bellevue, WA p. A URL: https://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/download/7085/6497
30. Zheng X, Han J, Sun A. A Survey of Location Prediction on Twitter. IEEE Trans Knowl Data Eng 2018 Sep 1;30(9):1652-1671. [doi: 10.1109/tkde.2018.2807840]
31. Han B, Cook P, Baldwin T. Text-Based Twitter User Geolocation Prediction. J Artif Intell Res 2014 Mar 20;49:451-500. [doi: 10.1613/jair.4200]
32. Pavalanathan U, Eisenstein J. Confounds and Consequences in Geotagged Twitter Data. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.: Association for Computational Linguistics; 2015 Presented at: 2015 Conference on Empirical Methods in Natural Language Processing; September 17-21, 2015; Lisbon, Portugal p. 2138-2148. [doi: 10.18653/v1/d15-1256]
33. Dredze M, Osborne M, Kambadur P. Geolocation for Twitter: Timing Matters. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.: Association for Computational Linguistics; 2016 Presented at: 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 12-17, 2016; San Diego, CA p. 1064-1069. [doi: 10.18653/v1/n16-1122]
34. Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C. Geo-located Twitter as proxy for global mobility patterns. Cartogr Geogr Inf Sci 2014 May 27;41(3):260-271. [doi: 10.1080/15230406.2014.890072] [Medline: 27019645]
35. Dredze M, García-Herranz M, Rutherford A, Mann G. Twitter as a Source of Global Mobility Patterns for Social Good. In: Proceedings of the 2016 ICML Workshop on #Data4Good: Machine Learning in Social Good Applications. In: ICML Workshop on #Data4Good: Machine Learning in Social Good Applications; 2016 Presented at: 2016 ICML Workshop on #Data4Good: Machine Learning in Social Good Applications; June 24, 2016; New York, NY.
36. Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D. Understanding Human Mobility from Twitter. PLoS One 2015 Jul 8;10(7):e0131469 [FREE Full text] [doi: 10.1371/journal.pone.0131469] [Medline: 26154597]
37. Huang Q, Wong D. Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data. Annals of the Association of American Geographers 2015 Sep 23;105(6):1179-1197. [doi: 10.1080/00045608.2015.1081120]
38. Birkin M, Harland K, Malleson N, Cross P, Clarke M. An Examination of Personal Mobility Patterns in Space and Time Using Twitter. International Journal of Agricultural and Environmental Information Systems (5) 2014:55-72. [doi: 10.4018/ijaeis.2014070104]
39. Hasan S, Zhan X, Ukkusuri S. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In: UrbComp '13: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing. 2013 Presented at: 2nd ACM SIGKDD International Workshop on Urban Computing; August 11, 2013; Chicago, IL p. 1-8. [doi: 10.1145/2505821.2505823]
40. Luo F, Cao G, Mulligan K, Li X. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. Applied Geography 2016 May;70:11-25. [doi: 10.1016/j.apgeog.2016.03.001]
41. Soliman A, Soltani K, Yin J, Padmanabhan A, Wang S. Social sensing of urban land use based on analysis of Twitter users' mobility patterns. PLoS One 2017 Jul 19;12(7):e0181657 [FREE Full text] [doi: 10.1371/journal.pone.0181657] [Medline: 28723936]
42. Kurkcu A, Ozbay K, Morgul E. Evaluating the Usability of Geo-located Twitter as a Tool for Human Activity and Mobility Patterns: A Case Study for New York City. 2016 Presented at: Transportation Research Board 95th Annual Meeting; January 10-14, 2016; Washington, DC.
43. Steiger E, Ellersiek T, Resch B, Zipf A. Uncovering Latent Mobility Patterns from Twitter During Mass Events. GI Forum 2015;1:525-534. [doi: 10.1553/giscience2015s525]

44. Twitter Social Mobility Index: A measure of social distancing derived from Twitter. URL: http://socialmobility.covid19dataresources.org/index [accessed 2020-10-28]
45. Filter realtime Tweets. Twitter Developer. URL: https://developer.twitter.com/en/docs/tweets/filter-realtime/overview/statuses-filter [accessed 2020-10-28]
46. geopy. github. URL: https://github.com/geopy/geopy [accessed 2020-10-28]
47. 15 Days to Slow the Spread. The White House. 2020 Mar 16. URL: https://www.whitehouse.gov/articles/15-days-slow-spread/ [accessed 2020-10-28]
48. Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. Signal Process 2020 Feb;167:107299. [doi: 10.1016/j.sigpro.2019.107299]
49. Lasry A, Kidder D, Hast M, Poovey J, Sunshine G, Winglee K, CDC Public Health Law Program, New York City Department of Health and Mental Hygiene, Louisiana Department of Health, San Francisco COVID-19 Response Team, Alameda County Public Health Department, San Mateo County Health Department, Marin County Division of Public Health. Timing of Community Mitigation and Changes in Reported COVID-19 and Community Mobility - Four U.S. Metropolitan Areas, February 26-April 1, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 17;69(15):451-457 [FREE Full text] [doi: 10.15585/mmwr.mm6915e2] [Medline: 32298245]
50. Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE. GitHub. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2020-10-28]
51. Raifman J, Nocka K, Jones D, Bor J, Lipson S, Jay J, et al. COVID-19 US State Policy Database. OpenICPSR. URL: https://www.openicpsr.org/openicpsr/project/119446/version/V39/view [accessed 2020-10-28]
52. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature 2020 Jun;582(7812):389-394. [doi: 10.1038/s41586-020-2284-y] [Medline: 32349120]
53. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J. Understanding the Demographics of Twitter Users. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2816/3234
54. Wojcik S, Hughes A. Sizing Up Twitter Users. Pew Research Center. 2019 Apr 24. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2020-10-28]
55. Zhao B, Sui D. True lies in geospatial big data: detecting location spoofing in social media. Annals of GIS 2017 Jan 31;23(1):1-14. [doi: 10.1080/19475683.2017.1280536]

Abbreviations

API: application programming interface
JSON: JavaScript Object Notation

Edited by G Eysenbach, G Fagherazzi; submitted 16.06.20; peer-reviewed by A Sun, J Li; comments to author 13.07.20; revised version received 04.08.20; accepted 11.10.20; published 03.12.20.

Please cite as:
Xu P, Dredze M, Broniatowski DA
The Twitter Social Mobility Index: Measuring Social Distancing Practices With Geolocated Tweets
J Med Internet Res 2020;22(12):e21499
URL: https://www.jmir.org/2020/12/e21499
doi:10.2196/21499
PMID:33048823

Original Paper

Tweets by People With Arthritis During the COVID-19 Pandemic: Content and Sentiment Analysis

Danielle Berkovic[1], BHSc (Hons); Ilana N Ackerman[1], BPhysio (Hons), PhD; Andrew M Briggs[2], BSc (Hons), PhD, FACP; Darshini Ayton[1], BBiomedSci (Hons), MPH, PhD

[1]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[2]School of Physiotherapy and Exercise Science, Curtin University, Perth, Australia

Corresponding Author:
Darshini Ayton, BBiomedSci (Hons), MPH, PhD
School of Public Health and Preventive Medicine
Monash University
Melbourne,
Australia
Phone: 61 425 705 130
Email: darshini.ayton@monash.edu

Abstract

Background: Emerging evidence suggests that people with arthritis are reporting increased physical pain and psychological distress during the COVID-19 pandemic. At the same time, Twitter's daily usage has surged by 23% throughout the pandemic period, presenting a unique opportunity to assess the content and sentiment of tweets. Individuals with arthritis use Twitter to communicate with peers, and to receive up-to-date information from health professionals and services about novel therapies and management techniques.

Objective: The aim of this research was to identify proxy topics of importance for individuals with arthritis during the COVID-19 pandemic, and to explore the emotional context of tweets by people with arthritis during the early phase of the pandemic.

Methods: From March 20 to April 20, 2020, publicly available tweets posted in English and with hashtag combinations related to arthritis and COVID-19 were extracted retrospectively from Twitter. Content analysis was used to identify common themes within tweets, and sentiment analysis was used to examine positive and negative emotions in themes to understand the COVID-19 experiences of people with arthritis.

Results: In total, 149 tweets were analyzed. The majority of tweeters were female and were from the United States. Tweeters reported a range of arthritis conditions, including rheumatoid arthritis, systemic lupus erythematosus, and psoriatic arthritis. Seven themes were identified: health care experiences, personal stories, links to relevant blogs, discussion of arthritis-related symptoms, advice sharing, messages of positivity, and stay-at-home messaging. Sentiment analysis demonstrated marked anxiety around medication shortages, increased physical symptom burden, and strong desire for trustworthy information and emotional connection.

Conclusions: Tweets by people with arthritis highlight the multitude of concurrent concerns during the COVID-19 pandemic. Understanding these concerns, which include heightened physical and psychological symptoms in the context of treatment misinformation, may assist clinicians to provide person-centered care during this time of great health uncertainty.

(J Med Internet Res 2020;22(12):e24550) doi:10.2196/24550

KEYWORDS

COVID-19; SARS-CoV-2; novel coronavirus; social media; Twitter; content analysis; sentiment analysis; microblogging; arthritis

Introduction

Social media contains a plethora of health information pertaining to individuals living with chronic illness [1,2]. Social media provides a unique opportunity to observe thoughts, feelings, and interactions between individuals living with chronic illness, and to leverage this information to promote positive health outcomes [3,4]. The COVID-19 pandemic has seen widespread uptake of social media use. Twitter, a well-known social media platform primarily used for microblogging, plays a significant role in crisis communications and can be a powerful tool to communicate to entire populations during a time of rapid change [5]. Twitter is already frequently used by individuals with arthritis to communicate with peers [6] and to receive up-to-date information from health professionals and services about novel therapies and management techniques [7].

Many inflammatory arthritis medications act as immunosuppressants, which are advantageous in controlling arthritis-mediated inflammatory responses, but can increase the risk of infection [8]. Long-term use of immune-modulating therapies or glucocorticoids may place individuals with arthritis in a higher-risk category for contracting the novel coronavirus SARS-CoV-2, although the level of risk is poorly understood [9-12]. Current guidelines suggest that individuals living with arthritis should physically distance from other individuals and their communities [13], and will likely need to do so for a longer duration than the general public. Heightened stress due to potential medication shortages, reduced opportunities to personally consult health care professionals, and enforced limitations on physical activity (which, for many, is a core component of arthritis self-management [14]) contribute to worsening arthritis symptoms, including disease flares. At present, individuals with arthritis have already reported increased physical pain and psychological distress during the COVID-19 pandemic [15,16].

In the current COVID-19 outbreak, Twitter's overall daily usage has surged by 23% in 2020 [17], presenting a unique opportunity to assess the content and sentiment of tweets. Examining publicly available tweets allows exploration of important proxy topics through microblogging data, without directly burdening this population. This research aims to identify proxy topics of importance for people with arthritis (of any diagnostic category) during the COVID-19 pandemic by characterizing the textual content and sentiment of tweets, and to explore the emotional context of tweets by people with arthritis during the early phase of the pandemic.

Methods

Design

An exploratory content and sentiment analysis was undertaken. All data were collected and reported according to the terms and conditions of Twitter, which state that content posted by individuals is publicly available to syndicate, broadcast, distribute, retweet, promote, or publish, excluding private information (eg, home addresses or identity documents) [18]. Use of tweets by individuals outside of Twitter can be carried out with no compensation paid to the individual tweeter, as use of Twitter is agreed upon as sufficient compensation [18]. The Monash University Human Research Ethics Committee (Project ID 24354) approved this project.

Inclusion and Exclusion Criteria

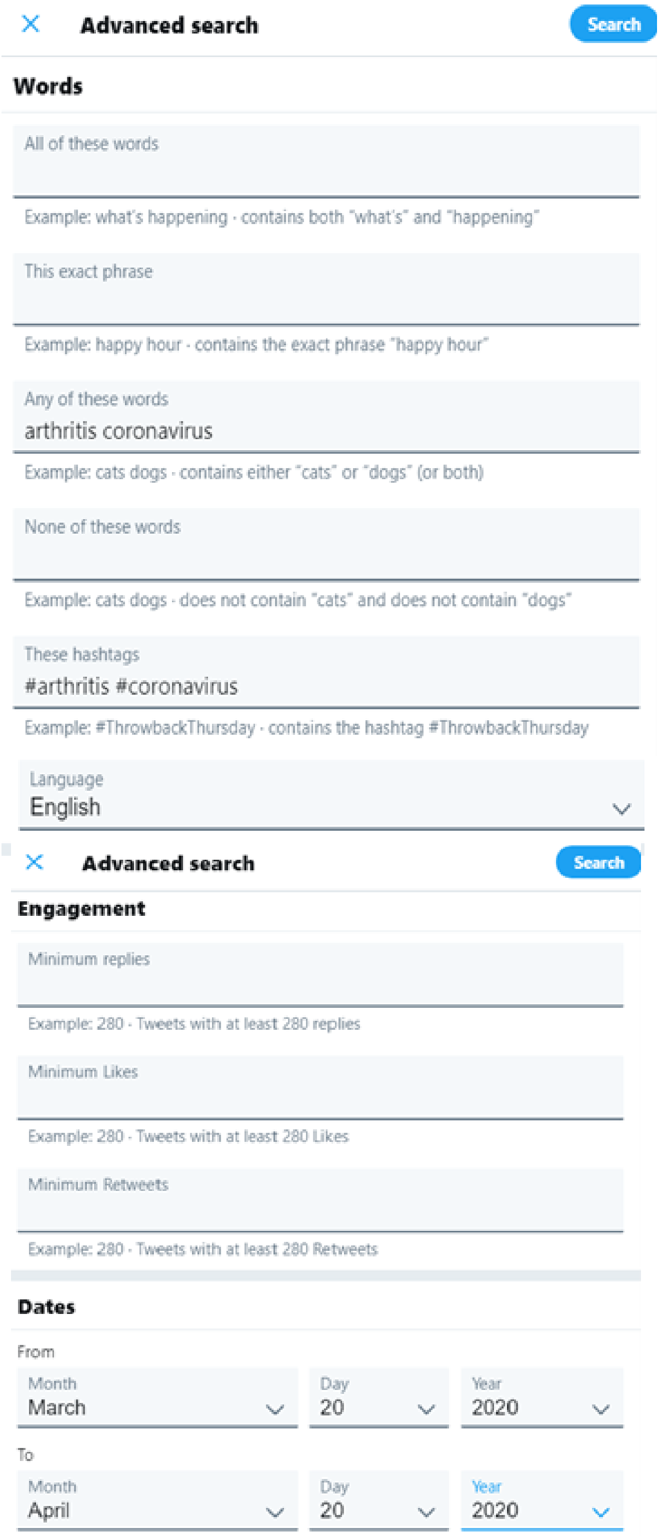
Publicly available tweets posted in English or with English translation (automated through Twitter), with the hashtags shown in Table 1 were included.

Table 1. Hashtags categorized by topic.

Topic	Hashtag
COVID-19	• #coronavirus • #covid19 • #isolation • #socialdistancing
Arthritis	• #arthritis • #spoonie • #rheumatologist • #rheumatology

Hashtags were selected based on trialing various combinations through Twitter's search function. The highest number of tweets retrieved were for the hashtags #coronavirus, #covid19, #arthritis, and #spoonie (the latter term coined by people living with chronic illness to describe various methods of pain management [7]). Hashtags were also searched as words; for example, where #arthritis was searched, arthritis without a hashtag was searched as well. This ensured that specific arthritis types mentioned in tweets without hashtags (eg, rheumatoid arthritis or psoriatic arthritis) were included. Hashtags and words were combined through the advanced search function on Twitter using the domains Hashtags, Words, Language, and Dates (Figure 1 [19]). Tweets were excluded if they originated from organizations, news outlets, or health professionals rather than individuals in order to focus on the personal perspective.

Figure 1. The advanced search function on Twitter [19].



Words

All of these words

Example: what's happening · contains both "what's" and "happening"

This exact phrase

Example: happy hour · contains the exact phrase "happy hour"

Any of these words: arthritis coronavirus

Example: cats dogs · contains either "cats" or "dogs" (or both)

None of these words

Example: cats dogs · does not contain "cats" and does not contain "dogs"

These hashtags: #arthritis #coronavirus

Example: #ThrowbackThursday · contains the hashtag #ThrowbackThursday

Language: English

Engagement

Minimum replies

Example: 280 · Tweets with at least 280 replies

Minimum Likes

Example: 280 · Tweets with at least 280 Likes

Minimum Retweets

Example: 280 · Tweets with at least 280 Retweets

Dates

From

Month: March
Day: 20
Year: 2020

To

Month: April
Day: 20
Year: 2020

Data Collection

Tweets were retrospectively extracted from March 20 to April 20, 2020. The search strategy and search results are included in Multimedia Appendix 1. The search timeframe was chosen to align with the early phase of the COVID-19 pandemic and the period when many developed countries (eg, the United States, the United Kingdom, Italy, Australia) announced enforceable physical distancing or isolation measures [20-23].

The desktop version of the Twitter website (versus the mobile app) was used for data collection for ethical purposes with only publicly available tweets extracted, rather than through a private login. In addition to the tweets themselves, accompanying data fields were extracted from each tweet using a customized template. Extracted data fields included (where possible): Twitter profile blurb, gender of tweeter, country of tweeter, number of likes, number of retweets, number of replies, hashtags used, number of hashtags, and use of accompanying photos.

Data were stored in a Microsoft Excel spreadsheet (v16.0, Microsoft Corp).

Data Analysis

To address the research aims, two data analysis techniques were utilized: summative content analysis and sentiment analysis. Content analysis was used to characterize the textual contents of tweets related to arthritis and COVID-19. Content analysis is exploratory; it aims to quantify and describe unknown phenomena [24]. During the content analysis process, the primary researcher (DB) read each tweet and categorized the tweets into a representative theme and subtheme related to a topic of importance for people with arthritis during the study period.

Summative content analysis occurred through a process of coding, which involved counting and comparisons of Twitter content, followed by interpretation of the underlying context [25]. To begin with, the first 10 tweets were analyzed and allocated a summary code. The code represents the theme of a tweet (eg, “health care experiences”). As additional tweets were examined, they were given one of the original codes or allocated a new code based on new content. This process was repeated until each tweet was coded and themed. Once these original themes were finalized, they were recoded for additional context, and a second researcher (DA) checked the coding. For example, “health care experiences” was coded for similarities in people’s health care experiences, such as difficulties accessing medications. Given Twitter’s character limits, each tweet only contained one theme. The frequency of original themes and subthemes was counted to indicate importance [26].

Sentiment analysis enables an examination of written and spoken words for positive and/or negative emotion. When applied to health care or social media research, sentiment analysis facilitates interpretation of textual information about patient experience from a person-centered perspective [27]. Once tweets were coded and categorized into themes, sentiment analysis was employed to assess the emotion associated with the theme using Glaser and Strauss’s [28] 6 codes for sentiment analysis, a common framework used for Twitter-based research [29,30]:

1. No sentiment: the tweet has no emotion or words or special punctuation; is matter-of-fact sounding;
2. Wretched: the tweet is purely negative;
3. Bad: the tweet contains mainly negative phrases and words that outweigh any positive sentiment;
4. So-so: the tweet has a mediocre and balanced sentiment where positive and negative statements are balanced;
5. Swell: the tweet contains mainly positive phrases and words which outweigh negative sentiment;
6. Great: the tweet is purely positive.

The presence of emojis, which are shorthand facial expression symbols that are frequently used to facilitate communication of mood and emotion, in tweets were also analyzed. To provide information regarding the emotional content of the tweets, Emoji Sentiment Ranking, as outlined by Kralj Novak et al [31], was applied. Tweets containing an emoji were categorized into one of three sentiment scores: (1) negative, (2) neutral, and (3) positive. Together, the content and sentiment analyses provide a proxy indicator of the topics of interest for, and perceived emotions of, people with arthritis during the COVID-19 pandemic.

Even in social media studies, it is imperative to protect participant anonymity [32]. To avoid reverse identification of participants based on their tweets (which can be found through internet searches), tweets analyzed in this study are not quoted verbatim. Instead, all data are expressed in aggregate form through descriptive statistics and qualitative syntheses.

Results

The analysis included 149 tweets posted during the study period. The majority of tweeters were female and based in the United States. The most common arthritis type was rheumatoid arthritis. Table 2 outlines gender, country of residence, and arthritis type.

Table 2. Demographics of Twitter users sampled.

Characteristic	Users (N=149), n (%)
Gender	
Female	105 (70.0)
Male	31 (21.0)
Unknown	13 (9.0)
Country	
United States	68 (45.5)
United Kingdom and Northern Ireland	39 (26)
Canada	18 (12.0)
Australia	1 (0.7)
France	1 (0.7)
Germany	1 (0.7)
India	1 (0.7)
New Zealand	1 (0.7)
Unknown	19 (13.0)
Diagnostic category	
Arthritis (specific arthritis type unclear)	86 (58.0)
Rheumatoid arthritis	32 (21.0)
Systemic lupus erythematosus	12 (8.0)
Psoriatic arthritis	10 (7.0)
Ankylosing spondylitis	3 (2.0)
Osteoarthritis	3 (2.0)
Juvenile idiopathic arthritis	3 (2.0)

Content analysis revealed seven themes from the tweets: (1) health care experiences, (2) personal stories, (3) links to or advertisements of relevant blogs, (4) discussion of arthritis-related symptoms, (5) advice sharing, (6) messages of positivity, and (7) stay-at-home messaging. Table 3 details the original themes and subthemes.

Table 3. Content analysis of themes and subthemes.

Theme and subthemes	Tweets, n (%)	Examples of phrases or #hashtags describing content
Health care experiences	55 (37.0)	• "I'm a long-term user of #hydroxychloroquine"
Difficulties accessing hydroxychloroquine	20 (36.5)	• "#Hydroxychloroquine destroyed my red blood cells"
Past experiences using hydroxychloroquine	20 (36.5)	
Support for President Trump's advice to use hydroxychloroquine to cure COVID-19	9 (16.0)	
Experiences within the National Health Service (UK NHS)	4 (7.0)	
Managing medication changes during COVID-19	2 (4.0)	
Personal stories	29 (20.0)	• "My rheumatologist has asked that I go into isolation. Now all I can do is enjoy the world from my window"
Explanation of history of managing arthritis, and subsequent fears of contracting or dying from COVID-19	9 (31.0)	• "My immune system is compromised but I'm being told to go to a germy hospital???"
Description of ways to self-manage physical symptoms (eg, exercising, staying connected with friends)	7 (24.0)	
Physical and psychological challenges of socially distancing	7 (24.0)	
Perceived barriers to attending rheumatologist appointments (eg, discomfort of wearing masks, fear of entering a high-risk location)	6 (21.0)	
Links to or advertisements of relevant blogs and forums	22 (14.0)	• "To our members, subscribers, followers, and fans: we are here for you. #BeSafe"
Recommendations and links from individuals to official patient- and consumer-led blogs (eg, CreakyJoints)	10 (45.0)	
Personal blogs on individual COVID-19 experiences (eg, how to manage worsening symptoms)	7 (32.0)	
Unofficial patient blogs (eg, online communities and forums) to create support networks for individuals	5 (23.0)	
Discussion of arthritis-related symptoms	15 (10.0)	• "Anyone else's arthritis flaring due to extra phone use?"
Increased physical pain	11 (73.5)	
Difficulty sleeping	2 (13.5)	
Reduced dexterity	2 (13.5)	
Advice seeking and sharing	14 (9.0)	• "Any suggestions for chronic pain sufferers? Coronavirus has been very rough #arthritis #spoonie"
Questions directed at government bodies (eg, the NHS and national working-from-home regulations)	5 (36.0)	• "Do we know if those of us with autoimmune conditions (arthritis) have a higher risk from #coronavirus"
Seeking advice from physiotherapists on at-home exercises to manage physical symptom burden	3 (22.0)	
Advice on whether to temporarily cease taking immunosuppressant medications	2 (14.0)	
Advice on how to protect airways if dexterity limitations prevent mask-wearing	2 (14.0)	
Questions directed to delivery services regarding delays	2 (14.0)	
Messages of positivity	8 (6.0)	• "It's amazing how motivating isolation can be! On my bike but knee sore #arthritis"
Gratitude for friends, family, and to still be able to appreciate life	4 (50.0)	
Spare time as a result of physical distancing facilitating more time to exercise and reduce physical symptom burden	4 (50.0)	
Stay-at-home messaging	6 (4.0)	• #stayathome
Emotional appeals for people to stay at home	4 (67.0)	• #arthritissucks
Angry appeals for people to stay at home	2 (33.0)	• #arthritiswarrior

The most common theme identified was experiences of navigating the health care system during the COVID-19 pandemic. Hydroxychloroquine (brand name Plaquenil) featured prominently in tweets, in terms of difficulties accessing the medication, past experiences using the medication, and recommendations from the President of the United States to use this medication to cure or prevent COVID-19, despite the lack of evidence or medical advice. Some individuals tweeted about their experiences within the National Health Service (NHS) in the United Kingdom, where patients were subject to longer-than-usual delays for rheumatology appointments, medication infusions, and general health check-ups.

Many individuals used Twitter as a platform to connect with and seek support from peers with arthritis. Tweets contained personal stories, links to personal and consumer-led blogs, discussion of arthritis-related symptoms, and advice seeking and sharing. Some individuals shared their challenges managing arthritis symptoms whilst being confined to their homes and questioned their physical and psychological capacity to function if they were to contract COVID-19. Others described the physical and emotional challenges associated with isolation, including increased physical pain, reduced dexterity, and missing family. Strategies to manage these symptoms included exercising and staying socially connected online with friends to ease mental strain. Tweeters were willing to guide others to potentially helpful resources, particularly blogs run by professional organizations (eg, CreakyJoints, the Arthritis Society [Canada]). Several tweets contained questions were directed toward government bodies (eg, regarding national working-from-home policies), whereas others reached out to physiotherapists or peers with arthritis for advice on appropriate exercises or lifestyle modifications to manage symptom burden during the isolation period. Some tweeters noted dexterity limitations that were highly relevant to COVID-19, such as being incapable of placing a mask behind their ears.

Some tweets were positive with tweeters noting they used their newfound spare time to concentrate on exercise, which was beneficial for mental health and pain reduction. Finally, some individuals expressed their desire for people to stay at home to flatten the curve of infections to return to normal life. Disapproval was voiced toward those refusing to practice physical distancing, whereas others expressed anger toward people not adhering to stay-at-home orders.

Sentiment analysis provided complementary information about the emotions associated with the content analysis themes. Table 4 details the original themes and corresponding sentiment, with phrases or hashtags describing tweeters' personal experiences.

Table 4. Sentiment analysis of tweets.

Original theme and sentiment[a]	Tweets, n (%)	Examples of phrases or #hashtags
Health care experiences	55 (37.0)	
Great	5 (9.0)	“doddle,” “virus gone in 3 days,” “NO SIDE EFFECTS,” “credible and good”
Swell	3 (5.0)	“thankful,” “it must be working,” “don’t be afraid”
So-so	8 (15.0)	“hope it helps,” “minimal risk,” “might work,” #BeWarnedBeWell, “it might fight COVID-19,” “take care world!” “interesting to see”
Bad	6 (11.0)	#plaquenilshortage,” “no tests will be done for 2 months,” “facing shortages,” “no proof #Hydroxychloroquine makes us safe”
Wretched	28 (51.0)	“seriously ill,” “madness,” “pissed off,” #coronavirushoax, #painsomnia, “f*ck all chance,” “made me sick,” “I could die,” “harmful results,” “side effects unbearable,” “I call bullsh*t,” “[medication] shortages & living with symptoms,” #EVIL!!, “shorten my life,” “life threatening for us,” “true danger,” “scared,” “fear a shortage,” #drugshortage, “will end up in hospital,” “write us all off,” “it made me so sick,” “NO real testing,” “it’s not safe”
No sentiment	5 (9.0)	“research says,” “UK government wants,” “prove hypothesis,” “IL-6 is raised,” “from the CDC”
Personal stories	29 (20.0)	
Great	5 (17.0)	“all good during this #covid19,” “it makes me feel better on a personal level,” “bring it on world,” “I am a champion!” #livingmybestlife
Swell	5 (17.0)	#AloneTogether, #nevergiveup, #hope, #makethebestofit, “trying to keep active”
So-so	3 (10.0)	“it is what it is,” “I hope they have a lot of masks and sanitizer!”
Bad	2 (7.5)	“I didn’t want to risk heading out,” “limiting my usual walk”
Wretched	12 (41.0)	“give us a break, FFS!” #HighRiskCovid19, “I hate #Coronavirus,” “literally a pain,” “holy sh*t,” “screaming into the void,” “I am not #expendable,” “being told to go to a germy hospital,” “my chances of surviving #covid19 are horrible,” “vulnerable patients like me”
No sentiment	2 (7.5)	“info is changing daily,” “I’m immunocompromised because”
Links to or advertisements of relevant blogs and forums	22 (14.0)	
Great	1 (5.0)	“so grateful”
Swell	2 (9.0)	“talking about work/life balance,” “we are here for you”
So-so	4 (18.0)	“coping in isolation,” “trying to deal,” “hoping to support others”
Bad	2 (9.0)	“how to handle flares,” “learn what’s happening to people with arthritis”
Wretched	4 (18.0)	“covid19 scariness,” “fear of dying,” “unpredictability and fear,” “we worry about everyone”
No sentiment	9 (41.0)	“video games during covid19,” “questions I have,” “share with your networks,” “breaking news,” “please consider sharing,” “please retweet,” “I found this information”
Discussion of arthritis-related symptoms	15 (10.0)	
Great	0 (0.0)	—[b]
Swell	0 (0.0)	—
So-so	1 (7.0)	“not the end of the world”
Bad	1 (7.0)	“I have to take a break”
Wretched	13 (86.0)	“struggling to sleep/be active,” “spoonie fail,” “the arthritis flared out of control,” “my shoulder is a casualty,” “pain in the knees,” “I’m screwed,” “ouchy grouchy,” “feet are burning,” “f*ck you coronavirus,” “worsening arthritis pain,” “I’m already achy,” “flaring due to extra phone use”
No sentiment	0 (0.0)	—
Advice seeking	14 (9.0)	
Great	0 (0.0)	—

Original theme and sentiment[a]	Tweets, n (%)	Examples of phrases or #hashtags
Swell	1 (7.0)	"thanks to the doctors for their expertise"
So-so	0 (0.0)	—
Bad	1 (7.0)	"some chance of developing complications"
Wretched	8 (57)	"lowered immunity more than it already is," "significantly increase risk of infection," "things have been rough," "zero immune system," "so stressed about #coronavirus," "no money for food let alone masks," "I'm disabled"
No sentiment	4 (29.0)	"grateful for info," "where can I volunteer to get tested," #lockdownUK, "would appreciate a video of exercises for in #lockdown"
Messages of positivity	8 (6.0)	
Great	2 (25.0)	"coronavirus caused some good things to happen," "amazing how motivating boredom is"
Swell	3 (37.5)	#selfmanagement, "right exercises to keep arthritis at bay," "enjoy the little things"
So-so	1 (12.5)	"life is too short to be scared"
Bad	1 (12.5)	"keeping positive a challenge"
Wretched	0 (0.0)	—
No sentiment	1 (12.5)	"this information might help"
Stay-at home-messaging	6 (4.0)	
Great	0 (0.0)	—
Swell	0 (0.0)	—
So-so	0 (0.0)	—
Bad	0 (0.0)	—
Wretched	5 (83.0)	"#selfisolation only because of my immune system," #arthritissucks so stay the f*ck home," "coronavirus could kill me #StayHomeSaveLives," "#PLEASESTAYHOME I'm devastated," "arthritis shot to sh*t stay home"
No sentiment	1 (17.0)	#GoHomeStayHome

[a] Great: the tweet is purely positive; swell: the tweet contains mainly positive phrases and words that outweigh negative sentiment; so-so: the tweet has a mediocre and balanced sentiment where positive and negative statements are balanced; bad: the tweet contains mainly negative phrases and words that outweigh any positive sentiment; wretched: the tweet is purely negative; no sentiment: the tweet has no emotion or words or special punctuation and is matter-of-fact sounding.

[b] Not available.

A few tweets contained messages of positivity. While overall "keeping positive [was] a challenge," some people encouraged others to "enjoy the little things" and that "life is too short to be scared." Still, the extent to which people with arthritis were concerned for their health was evident in people's stay-at-home messaging. Tweeters were notably anxious and angry, writing that "#arthritissucks so stay the f*ck home," that "coronavirus could kill me #StayHomeSaveLives," and "#PLEASESTAYHOME I'm devastated."

Using Glaser and Strauss's [28] classifications, more than half of the tweets contained wretched (purely negative) or bad (mainly negative) sentiment (n=83, 56%), whereas only one-fifth of tweets contained great (purely positive) or swell (mainly positive) sentiment (n=27, 18%). In total, 16 (11%) tweets contained sentiment that was so-so (balanced negativity and positivity), and 22 (15%) tweets contained no sentiment (matter-of-fact sounding).

Individuals in the United States appeared particularly despondent regarding their health care experiences during the COVID-19 pandemic. When referencing interactions with the health care system, tweets contained phrases such as "true danger," "seriously pissed off," and "scared." Tweeters noted "it's not safe" for people with arthritis facing hydroxychloroquine shortages, and that the #drugshortage was "life threatening." Some individuals mused that there is "minimal risk" trying hydroxychloroquine to cure COVID-19, and that it might be "interesting to see" the results of this medication. Outside of the United States, Canadians tweeted that engaging with the health care system during the pandemic was a "doddle" and that they were "thankful" to continue to have access to their health professionals.

Many individuals described their personal stories negatively. People with arthritis discussed barriers to accessing care, such as being surprised at "being told to go to a germy hospital" and that they "didn't want to risk going out." Tweets with links to blogs were accompanied by captions "fear of dying," "unpredictability and fear," and "we worry about everyone." Some were more positive, encouraging followers that they are "hoping to support others." Although their representation was small, people with arthritis in Ireland and New Zealand viewed the isolation measures as an opportunity for #hope, encouraging others to #makethebestofit and to #nevergiveup.

A consistent sentiment was that people with arthritis were negative in their discussion of physical and psychological symptoms, with many individuals seeking advice from peers and health professionals to remedy symptoms. Tweets highlighted the spectrum of symptoms that individuals experienced during the pandemic, with people “struggling to sleep/be active,” experiencing “worsening arthritis pain,” and having “pain in the knees.” Some noted that attempting to manage physical symptoms while in isolation was “literally a pain” and that the psychological toll was like “screaming into the void.” One tweeter mentioned that their arthritis was “flaring due to extra phone use,” which was frustrating since this was a primary method of maintaining social connection and communication with family or colleagues during the pandemic. Similarly, individuals tweeting and asking for advice did so by prefacing that “things have been rough,” that they have “zero immune system,” and that they are “so stressed about #coronavirus.”

Our sentiment analysis included an overview of emoji use in tweets, as summarized in Table 5.

Table 5. Sentiment analysis of emoji use in tweets.

Emoji[a]	Count, n	Emoji name	Phrases or #hashtags accompanying the emoji	Emotion	Link to original themes
	3	Face with no good gesture	“my feet are burning,” things are very painful,” “my chances of surviving are horrible”	Negative	Discussion of arthritis-related symptoms, personal stories
	2	Clapping hands sign	“my team have worked so hard,” “massive thanks”	Positive	Links to or advertisements of relevant blogs and forums, advice seeking
	2	Smiling face with smiley eyes	“hope it helps all”	Positive	Health care experiences
	2	Confused face	“I might be screwed,” “I think my consultant is wrong”	Negative	Health care experiences, discussion of arthritis-related symptoms
	2	Person with folded hands	#lupuswarrior	Positive	Personal stories, health care experiences
	2	Red heart suit	“hoping to support others,” #BeSafe	Positive	Links to or advertisements of relevant blogs and forums
	2	Purple heart	“choose to spread awareness”	Positive	Personal stories
	1	Crying face	“#coronavirus has opened my eyes”	Neutral	Health care experiences
	1	Smirking face	“#makingthebestofit	Positive	Personal stories
	1	Fisted hand sign	“I’ve got a morphine patch on to help”	Positive	Personal stories
	1	Face with tears of joy	“happy sitting in the sun”	Positive	Messages of positivity
	1	Leaves fluttering in the wind	“I’m laying outside”	Positive	Personal stories
	1	Face with mouth open	“could potentially lead to the cure for #Coronavirus”	Positive	Health care experiences
	1	Winking face	“I made healthy Easter treats in self isolation”	Positive	Personal stories
	1	Pill	“I really miss anti-inflammatories”	Negative	Stay-at-home messaging
	1	Front-facing baby chick	#livingmybestlife	Positive	Personal stories
	1 each	See-no-evil monkey, speak-no-evil monkey, hear-no-evil monkey, monkey, monkey face	“my rheumatologist said I wasn’t at high risk”	Positive	Health care experiences

[a] Emoji Sentiment Ranking did not include the upside-down face, the spectacle face, or the high-five emoji. There was one of each included in our data collection.

The emoji within tweets established tone and emphasized emotions. Most tweets used an emoji to express positive sentiment, such as clapping hands to thank colleagues for their support and a person with folded hands in acknowledgment of the personal commitment to the ongoing management of arthritis symptoms in isolation. The main emoji used to express negative sentiment were the face with “no good” gesture, and the confused face. Tweeters used the face with “no good” gesture to highlight their physical pain, and the confused face was used to express confusion around what arthritis patients perceived to be ill-informed health advice during the pandemic.

Discussion

Principal Findings

This study aimed to identify proxy topics of importance for people with arthritis during COVID-19 by characterizing the textual content and sentiment of tweets, and exploring the emotional context of tweets by people with arthritis during the early phase of the pandemic. We anticipated that this novel approach would ascertain contemporary topics of importance for immunocompromised and isolated individuals. Content analysis revealed seven themes relating to health care experiences, and sentiment analysis revealed that the majority of tweets contained negative emotion, particularly around medication shortages, increased arthritis symptoms, and the physical and mental toll of physical distancing or living in isolation. Our findings provide a starting point for understanding the impacts of COVID-19 on vulnerable arthritis populations and provide some insights for physicians and researchers regarding current concerns that may inform tailored care.

More than one-third of tweets discussed health care experiences, primarily focusing on the reduced availability of hydroxychloroquine. Individuals were highly anxious about hydroxychloroquine shortages after preliminary research found that the drug might act as a potential preventive measure, or possible cure, for COVID-19 and these early findings were highly publicized around the world. The original published article has since been retracted after researchers were unable to verify the reliability of their results [33]. Regardless, President Trump tweeted his support for hydroxychloroquine as “one of the biggest game changers in the history of medicine” [34], causing panic about medication shortages, particularly from those with systemic lupus erythematosus for which it is the first-line therapy [35]. Due to off-label prescriptions and hoarding practices, difficulties accessing hydroxychloroquine have been reported globally [36]. Hydroxychloroquine shortages pose a threat to the health and safety of people with inflammatory arthritis, with reports that many will experience flare-ups and may develop irreversible organ damage without their regular dose [35]. Our findings highlight the need for accurate information about treatments and their effectiveness and the critical role that clinicians play in dispelling myths and inaccuracies during times of rapidly changing information. A growing body of literature describes the potential benefits of using Twitter in clinical settings, and reports ways that clinicians are using platforms such as Twitter to communicate health information to the broader population [37-39]. In the context of infectious diseases in particular, evidence suggests that Twitter is beneficial to translate real-time clinical information [40,41]. Use of Twitter, being a component of mobile health (mHealth), also empowers patients to more positively perceive their abilities to manage chronic illness [42]. This presents an opportunity for clinicians and professional societies to use social media platforms such as Twitter to overcome evidence dissemination methods (eg, peer-reviewed articles and care guidelines) that are traditionally slow. Together, clinicians and patients can contribute to care in adults with arthritis, encouraging positive health outcomes throughout, and beyond, the pandemic.

Several tweets contained very personal narratives that highlighted individual fears of contracting COVID-19, and the challenges associated with being vulnerable to infection. Individuals described perceived barriers to accessing care, citing discomfort caused by wearing medical-grade personal protective equipment. Tweets contained accounts of neck pain attributed to wearing a mask; a documented side effect in other vulnerable populations [43]. Fitting a mask is also a dexterous task that some people with arthritis struggled to perform, and as a result, encountered abuse from others when buying groceries or going for a walk. The long-term psychosocial impact of this stigma should receive consideration in future research. Clinicians should consider COVID-19–related functional concerns for physically impaired patients, and incorporate new aspects of information-seeking into clinical consultations. For example, asking patients about their degree of difficulty with COVID-19–related functional tasks would help to elicit relevant functional challenges faced by people with arthritis and inform the provision of tailored patient education. Asking these questions also contributes to factors beyond patient education and care, such as facilitating access to nonacute symptom or pain management services such as allied health. These health services still need to be maintained during and after the pandemic, and for musculoskeletal health in particular, there is emerging commentary around the physical and psychosocial impacts of inhibited access [44,45].

Tweets detailed marked increases in general physical symptom burden; a concerning prospect as COVID-19 has impacted face to face consultations, with indications that disruptions to traditional service models will likely persist for some time. Evidence about the utility of telemedicine to manage pain is emerging, with consideration of barriers to implementation, and potential inequity in access [46], although health systems have been generally slow to implement this approach at scale. While the included tweets provide preliminary information about the growing symptom burden during the pandemic, the collection of systematic patient-reported outcomes data is needed to ensure that health care services are meeting the needs of people with arthritis during and after the COVID-19 pandemic.

A proportion of tweets related to social connection, that is, people reaching out to likeminded peers with arthritis through potentially informative blogs. Most links provided were to official (eg, CreakyJoints) or unofficial (eg, online communities and forums) blogs and provided information on how to manage physical and mental health in isolation, a range of arthritis-appropriate exercises, and existing evidence on the association between arthritis and risk of COVID-19 infection. Access to resources that are relevant, credible, and trustworthy appears to have been challenging for people with arthritis throughout the pandemic [47], and combined with high levels of misinformation online [48,49], this may account for the recent growth of platforms such as Twitter for sourcing information and advice.

Before the COVID-19 pandemic, people with arthritis primarily used social media for self-expression and positive messaging [50,51]. Our sentiment analysis (enhanced by classifying emojis to further characterize common emotions) demonstrates that the role of Twitter has evolved throughout the current pandemic to act as a space for people to share symptoms; to reach out to peers, organizations, and health professionals for information; and to create a sympathetic community of care. This is advantageous as it fosters connection between individuals with shared experiences but conversely may enable proliferation of misinformation [52]. Already, Twitter has been shown to inform clinical practice by capturing the experiences of patients with multiple sclerosis during the pandemic [53]. Understanding the COVID-19–related concerns of people with arthritis is also key to providing person-centered care and reducing distress during these rapidly changing times.

Strengths and Limitations

The observational exploratory nature of this study enabled us to examine topics of importance for individuals with arthritis through a person-centered lens, without ethical issues or compromising the well-being of immunocompromised patients during the pandemic. Social media research is still in its infancy, and this novel method of data collection demonstrates the concerns of people with arthritis during a time of peak anxiety. There is some indication that tweeters were representative of the general inflammatory arthritis population; the majority were female and the most common arthritis type identified was rheumatoid arthritis. Nearly half (44.0%) of tweeters were based in the United States, which currently leads the world in COVID-19 cases and deaths [54].

We also acknowledge the research limitations. It is important to note that only 15% of adults regularly use Twitter, and that younger adults and minority communities tend to be more highly represented on Twitter than the general population [55], although minority communities have been significantly impacted by COVID-19 [56,57]. Our results should therefore be interpreted as representing a small subset of people with arthritis, and not all people with the disease. Data were extracted rather than collected directly from people with arthritis and critics of social media research purport that posts or tweets are often curated and may not be reflective of reality. We have attempted to minimize this potential bias by conducting sentiment analysis, which helped us determine the emotional tone associated with Twitter content. Regardless, sentiment analysis has its limitations: populations and individuals are constantly stimulated by their political and socioeconomic surroundings and individual demographics, which can influence the content and sentiment of people's tweets [58,59]. We also recognize the potential limitations of our search strategy (eg, we did not search for hashtags related to specific symptoms, such as pain or function, that are not unique to arthritis) that may have impacted the number of retrieved tweets. While we had a modest sample size of tweets due to our focused study aims, sentiment analysis has previously been conducted in studies with comparable sample sizes of tweets (n=260 and n=200) [60,61]. We were unable to determine the specific diagnostic category for over half the people tweeting; while a small number of tweets were from people with osteoarthritis, it is possible that more may be represented within the "diagnosis not specified" category. We were only able to analyze tweets in English, and these largely came from high-income, developed countries. Tweets in other languages and those from people in low-and-middle income countries may provide further insights, especially where the prevalence of COVID-19 infection is high [62].

Conclusion

This study highlights the spectrum of concerns facing people with arthritis during the COVID-19 pandemic. By exploring the content and sentiment of recent tweets, we found that individuals with arthritis conditions experience marked anxiety about medication shortages and increased physical symptom burden, and are seeking connection with and information from peers. These findings can be used to raise awareness of key issues relevant to people with arthritis during the pandemic, and to guide clinicians to tailor care that addresses the specific concerns and needs of their patients during the pandemic.

Acknowledgments

DB received a PhD scholarship from Musculoskeletal Australia to conduct this research [PURE ID #230581862]. INA is supported by a Victorian Health and Medical Research Fellowship from the Victorian Government.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Twitter search strategy and results.
[DOCX File , 15 KB - jmir_v22i12e24550_app1.docx]

References

1. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, et al. Social media use among patients and caregivers: a scoping review. BMJ Open 2013 May 09;3(5):e002819 [FREE Full text] [doi: 10.1136/bmjopen-2013-002819] [Medline: 23667163]
2. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec 17;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
3. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013 Dec;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
4. Santoro E, Castelnuovo G, Zoppis I, Mauri G, Sicurello F. Social media and mobile applications in chronic disease prevention and management. Front Psychol 2015 May 07;6:567 [FREE Full text] [doi: 10.3389/fpsyg.2015.00567] [Medline: 25999884]

5. Eriksson M, Olsson E. Facebook and Twitter in Crisis Communication: A Comparative Study of Crisis Communication Professionals and Citizens. J Contingencies & Crisis Man 2016 Jun 13;24(4):198-208. [doi: 10.1111/1468-5973.12116]
6. Curtis JR, Chen L, Higginbotham P, Nowell WB, Gal-Levy R, Willig J, et al. Social media for arthritis-related comparative effectiveness and safety research and the impact of direct-to-consumer advertising. Arthritis Res Ther 2017 Mar 07;19(1):48 [FREE Full text] [doi: 10.1186/s13075-017-1251-y] [Medline: 28270190]
7. Kloth YM, Deutsch KM, Danielson KA, Strack J, Law C. What Twitter teaches us about patient-provider communication on pain. PLoS One 2019 Dec 26;14(12):e0226321 [FREE Full text] [doi: 10.1371/journal.pone.0226321] [Medline: 31877158]
8. Caporali R, Caprioli M, Bobbio-Pallavicini F, Montecucco C. DMARDS and infections in rheumatoid arthritis. Autoimmun Rev 2008 Dec;8(2):139-143. [doi: 10.1016/j.autrev.2008.05.001] [Medline: 19014871]
9. Coronavirus (COVID-19) advice for people with chronic conditions. Australian Government. 2020. URL: https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/advice-for-people-at-risk-of-coronavirus-covid-19/coronavirus-covid-19-advice-for-people-with-chronic-health-conditions [accessed 2020-09-10]
10. Fredi M, Cavazzana I, Moschetti L, Andreoli L, Franceschini F, Airò P, et al. COVID-19 in patients with rheumatic diseases in northern Italy: a single-centre observational and case–control study. The Lancet Rheumatology 2020 Sep;2(9):e549-e556. [doi: 10.1016/s2665-9913(20)30169-7]
11. Zhong J, Shen G, Yang H, Huang A, Chen X, Dong L, et al. COVID-19 in patients with rheumatic disease in Hubei province, China: a multicentre retrospective observational study. The Lancet Rheumatology 2020 Sep;2(9):e557-e564. [doi: 10.1016/s2665-9913(20)30227-7]
12. Gianfrancesco M, Hyrich KL, Al-Adely S, Carmona L, Danila MI, Gossec L, COVID-19 Global Rheumatology Alliance. Characteristics associated with hospitalisation for COVID-19 in people with rheumatic disease: data from the COVID-19 Global Rheumatology Alliance physician-reported registry. Ann Rheum Dis 2020 Jul 12;79(7):859-866 [FREE Full text] [doi: 10.1136/annrheumdis-2020-217871] [Medline: 32471903]
13. UPDATED: Advice regarding Coronavirus (COVID-19) from the Australian Rheumatology Association. Arthritis Australia. 2020. URL: https://arthritisaustralia.com.au/advice-regarding-coronavirus-covid-19-from-the-australian-rheumatology-association/ [accessed 2020-09-10]
14. Décary S, Barton JL, Proulx L, Richards DP, Paterson G, de Wit M, et al. How to Effectively Support Patients with Rheumatic Conditions Now and Beyond COVID-19. ACR Open Rheumatol 2020 Sep 28;2(9):505-506 [FREE Full text] [doi: 10.1002/acr2.11152] [Medline: 32403182]
15. Michaud K, Wipfler K, Shaw Y, Simon TA, Cornish A, England BR, et al. Experiences of Patients With Rheumatic Diseases in the United States During Early Days of the COVID-19 Pandemic. ACR Open Rheumatol 2020 Jun 09;2(6):335-343 [FREE Full text] [doi: 10.1002/acr2.11148] [Medline: 32311836]
16. Mobasheri A. COVID-19, osteoarthritis and women's health. Case Reports in Women's Health 2020 Jul;27:e00207. [doi: 10.1016/j.crwh.2020.e00207]
17. Gadde V, Derella M. An update on our continuity strategy during COVID-19. Twitter. 2020. URL: https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html [accessed 2020-09-10]
18. Terms of Service. Twitter. 2020. URL: https://twitter.com/en/tos [accessed 2020-09-10]
19. Advanced search function [image]. Twitter. URL: https://twitter.com/search-advanced?lang=en [accessed 2020-09-10]
20. Staying at home and away from others (social distancing). Gov.UK. 2020. URL: https://www.gov.uk/government/publications/full-guidance-on-staying-at-home-and-away-from-others [accessed 2020-09-10]
21. Update on Coronavirus Measures. Prime Minister of Australia. 2020. URL: https://www.pm.gov.au/media/update-coronavirus-measures-24-March-2020 [accessed 2020-09-10]
22. Vinceti M, Filippini T, Rothman KJ, Ferrari F, Goffi A, Maffeis G, et al. Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking. EClinicalMedicine 2020 Aug;25:100457 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100457] [Medline: 32838234]
23. Zhang R, Li Y, Zhang AL, Wang Y, Molina MJ. Identifying airborne transmission as the dominant route for the spread of COVID-19. Proc Natl Acad Sci U S A 2020 Jun 30;117(26):14857-14863 [FREE Full text] [doi: 10.1073/pnas.2009637117] [Medline: 32527856]
24. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep 11;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
25. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
26. Briggs A, Houlding E, Hinman R, Desmond L, Bennell K, Darlow B, et al. Health professionals and students encounter multi-level barriers to implementing high-value osteoarthritis care: a multi-national study. Osteoarthritis Cartilage 2019 May;27(5):788-804. [doi: 10.1016/j.joca.2018.12.024] [Medline: 30668988]
27. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013 Nov 01;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]

28. Glaser B, Strauss A. The Discovery of Grounded Theory: Strategies for Qualitative Research. London and New York: Routledge Taylor & Francis Group; 1967.
29. Jansen BJ, Zhang M, Sobel K, Chowdury A. Twitter power: Tweets as electronic word of mouth. J Am Soc Inf Sci 2009 Nov;60(11):2169-2188. [doi: 10.1002/asi.21149]
30. Williams SA, Terras MM, Warwick C. What do people study when they study Twitter? Classifying Twitter related academic papers. Journal of Documentation 2013 May 10;69(3):384-410. [doi: 10.1108/jd-03-2012-0027]
31. Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of Emojis. PLoS One 2015 Dec 7;10(12):e0144296 [FREE Full text] [doi: 10.1371/journal.pone.0144296] [Medline: 26641093]
32. Ayers JW, Caputi TL, Nebeker C, Dredze M. Don't quote me: reverse identification of research participants in social media studies. NPJ Digit Med 2018 Aug 2;1(1):30 [FREE Full text] [doi: 10.1038/s41746-018-0036-2] [Medline: 31304312]
33. Mehra MR, Desai SS, Ruschitzka F, Patel AN. RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. The Lancet 2020 May. [doi: 10.1016/s0140-6736(20)31180-6]
34. @realDonaldTrump. Twitter. 2020 Mar 22. URL: https://twitter.com/realDonaldTrump/status/1241367239900778501 [accessed 2020-09-10]
35. Mehta B, Salmon J, Ibrahim S. Potential Shortages of Hydroxychloroquine for Patients with Lupus During the Coronavirus Disease 2019 Pandemic. JAMA Health Forum 2020 Apr 10;1(4):e200438. [doi: 10.1001/jamahealthforum.2020.0438]
36. Choo EK, Rajkumar SV. Medication Shortages During the COVID-19 Crisis: What We Must Do. Mayo Clin Proc 2020 Jun;95(6):1112-1115 [FREE Full text] [doi: 10.1016/j.mayocp.2020.04.001] [Medline: 32312491]
37. Pershad Y, Hangge P, Albadawi H, Oklu R. Social Medicine: Twitter in Healthcare. J Clin Med 2018 May 28;7(6):121 [FREE Full text] [doi: 10.3390/jcm7060121] [Medline: 29843360]
38. Hawn C. Take Two Aspirin And Tweet Me In The Morning: How Twitter, Facebook, And Other Social Media Are Reshaping Health Care. Health Affairs 2009 Mar;28(2):361-368. [doi: 10.1377/hlthaff.28.2.361]
39. Cifu AS, Vandross AL, Prasad V. Case Reports in the Age of Twitter. Am J Med 2019 Oct;132(10):e725-e726. [doi: 10.1016/j.amjmed.2019.03.044] [Medline: 30998916]
40. Goff DA, Kullar R, Newland JG. Review of Twitter for infectious diseases clinicians: useful or a waste of time? Clin Infect Dis 2015 May 15;60(10):1533-1540. [doi: 10.1093/cid/civ071] [Medline: 25652087]
41. Ghosh P, Schwartz G, Narouze S. Twitter as a powerful tool for communication between pain physicians during COVID-19 pandemic. Reg Anesth Pain Med 2020 Apr 21. [doi: 10.1136/rapm-2020-101530] [Medline: 32321859]
42. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End User and Implementer Experiences of mHealth Technologies for Noncommunicable Chronic Disease Management in Young Adults: Systematic Review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
43. Ong JJ, Bharatendu C, Goh Y, Tang JZ, Sooi KW, Tan YL, et al. Headaches Associated With Personal Protective Equipment - A Cross-Sectional Study Among Frontline Healthcare Workers During COVID-19. Headache 2020 May 12;60(5):864-877. [doi: 10.1111/head.13811] [Medline: 32232837]
44. Karos K, McParland JL, Bunzli S, Devan H, Hirsh A, Kapos FP, et al. The social threats of COVID-19 for people with chronic pain. Pain 2020 Oct 13;161(10):2229-2235 [FREE Full text] [doi: 10.1097/j.pain.0000000000002004] [Medline: 32694381]
45. Nestola T, Orlandini L, Beard JR, Cesari M. COVID-19 and Intrinsic Capacity. J Nutr Health Aging 2020 May 28;24(7):692-695. [doi: 10.1007/s12603-020-1397-1] [Medline: 32744562]
46. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020 May;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
47. Duron G, Gelman L, Dua A, Putman M. Tracking clinical resources for coronavirus disease 2019. Curr Opin Rheumatol 2020 Sep;32(5):441-448. [doi: 10.1097/BOR.0000000000000724] [Medline: 32675716]
48. Li HO, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health 2020 May 14;5(5):e002604 [FREE Full text] [doi: 10.1136/bmjgh-2020-002604] [Medline: 32409327]
49. Erku DA, Belachew SA, Abrha S, Sinnollareddy M, Thomas J, Steadman KJ, et al. When fear and misinformation go viral: Pharmacists' role in deterring medication misinformation during the 'infodemic' surrounding COVID-19. Res Social Adm Pharm 2020 May 01 [FREE Full text] [doi: 10.1016/j.sapharm.2020.04.032] [Medline: 32387230]
50. Modica R, Lomax KG, Batzel P, Cassanas A. OARRR 2018 Jun;Volume 10:73-81. [doi: 10.2147/oarrr.s165010]
51. des Bordes JKA, Foreman J, Westrich-Robertson T, Lopez-Olivo MA, Peterson SK, Hofstetter C, et al. Interactions and perceptions of patients with rheumatoid arthritis participating in an online support group. Clin Rheumatol 2020 Jun 31;39(6):1775-1782. [doi: 10.1007/s10067-020-04967-y] [Medline: 32006180]
52. Malecki K, Keating J, Safdar N. Crisis Communication and Public Perception of COVID-19 Risk in the Era of Social Media. Clin Infect Dis 2020 Jun 16:2020 [FREE Full text] [doi: 10.1093/cid/ciaa758] [Medline: 32544242]
53. Nesbitt C, Rath L, Yeh WZ, Zhong M, Wesselingh R, Monif M, et al. MSCOVID19: Using social media to achieve rapid dissemination of health information. Mult Scler Relat Disord 2020 Oct;45:102338. [doi: 10.1016/j.msard.2020.102338] [Medline: 32629402]

54. Schneider EC. Failing the Test — The Tragic Data Gap Undermining the U.S. Pandemic Response. N Engl J Med 2020 Jul 23;383(4):299-302. [doi: 10.1056/nejmp2014836]
55. Smith A, Brenner R. Twitter Use 2012. Pew Research Center Internet & Technology. 2012 May 31. URL: https://www.pewresearch.org/internet/2012/05/31/twitter-use-2012/ [accessed 2020-09-10]
56. Clark E, Fredricks K, Woc-Colburn L, Bottazzi ME, Weatherhead J. Disproportionate impact of the COVID-19 pandemic on immigrant communities in the United States. PLoS Negl Trop Dis 2020 Jul 13;14(7):e0008484 [FREE Full text] [doi: 10.1371/journal.pntd.0008484] [Medline: 32658925]
57. Pan D, Sze S, Minhas JS, Bangash MN, Pareek N, Divall P, et al. The impact of ethnicity on clinical outcomes in COVID-19: A systematic review. EClinicalMedicine 2020 Jun;23:100404 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100404] [Medline: 32632416]
58. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
59. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLoS One 2015 Sep 2;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
60. Nemes L, Kiss A. Social media sentiment analysis based on COVID-19. Journal of Information and Telecommunication 2020 Jul 14:1-15. [doi: 10.1080/24751839.2020.1790793]
61. Clyne W, Pezaro S, Deeny K, Kneafsey R. Using Social Media to Generate and Collect Primary Data: The #ShowsWorkplaceCompassion Twitter Research Campaign. JMIR Public Health Surveill 2018 Apr 23;4(2):e41 [FREE Full text] [doi: 10.2196/publichealth.7686] [Medline: 29685866]
62. Walker P, Whittaker C, Watson OJ, Baguelin M, Winskill P, Hamlet A, et al. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science 2020 Jul 24;369(6502):413-422 [FREE Full text] [doi: 10.1126/science.abc0035] [Medline: 32532802]

Abbreviations

mHealth: mobile health
NHS: National Health Service

Edited by G Eysenbach; submitted 24.09.20; peer-reviewed by R Safari, R Gore; comments to author 09.10.20; revised version received 26.10.20; accepted 28.10.20; published 03.12.20.

Please cite as:
Berkovic D, Ackerman IN, Briggs AM, Ayton D
Tweets by People With Arthritis During the COVID-19 Pandemic: Content and Sentiment Analysis
J Med Internet Res 2020;22(12):e24550
URL: https://www.jmir.org/2020/12/e24550
doi:10.2196/24550
PMID:33170802

Original Paper

Dynamic Public Health Surveillance to Track and Mitigate the US COVID-19 Epidemic: Longitudinal Trend Analysis Study

Lori Ann Post[1], PhD; Tariq Ziad Issa[2], BA; Michael J Boctor[2], BSc; Charles B Moss[3], PhD; Robert L Murphy[4], MD; Michael G Ison[5], MSc, MD; Chad J Achenbach[5], MPH, MD; Danielle Resnick[6], PhD; Lauren Nadya Singh[1], MPH; Janine White[1], MSc; Joshua Marco Mitchell Faber[2], BA; Kasen Culler[2], BSc; Cynthia A Brandt[7], MPH, MD; James Francis Oehmke[1], PhD

[1]Buehler Center for Health Policy & Economics and Departments of Emergency Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, United States

[4]Center for Global Communicable Diseases, Institute for Global Health, Northwestern University, Chicago, IL, United States

[5]Divsion of Infectious Disease, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[6]International Food Policy Research Institute, Washington, DC, United States

[7]Yale Center for Medical Informatics, Yale School of Medicine, Yale University, New Haven, CT, United States

Corresponding Author:
Lori Ann Post, PhD
Buehler Center for Health Policy & Economics and Departments of Emergency Medicine
Feinberg School of Medicine
Northwestern University
750 N Lake Shore Dr
9th Floor, Suite 9-9035 Rubloff Building
Chicago, IL, 60611
United States
Phone: 1 203 980 7107
Email: lori.post@northwestern.edu

Abstract

Background: The emergence of SARS-CoV-2, the virus that causes COVID-19, has led to a global pandemic. The United States has been severely affected, accounting for the most COVID-19 cases and deaths worldwide. Without a coordinated national public health plan informed by surveillance with actionable metrics, the United States has been ineffective at preventing and mitigating the escalating COVID-19 pandemic. Existing surveillance has incomplete ascertainment and is limited by the use of standard surveillance metrics. Although many COVID-19 data sources track infection rates, informing prevention requires capturing the relevant dynamics of the pandemic.

Objective: The aim of this study is to develop dynamic metrics for public health surveillance that can inform worldwide COVID-19 prevention efforts. Advanced surveillance techniques are essential to inform public health decision making and to identify where and when corrective action is required to prevent outbreaks.

Methods: Using a longitudinal trend analysis study design, we extracted COVID-19 data from global public health registries. We used an empirical difference equation to measure daily case numbers for our use case in 50 US states and the District of Colombia as a function of the prior number of cases, the level of testing, and weekly shift variables based on a dynamic panel model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in R.

Results: Examination of the United States and state data demonstrated that most US states are experiencing outbreaks as measured by these new metrics of speed, acceleration, jerk, and persistence. Larger US states have high COVID-19 caseloads as a function of population size, density, and deficits in adherence to public health guidelines early in the epidemic, and other states have alarming rates of speed, acceleration, jerk, and 7-day persistence in novel infections. North and South Dakota have had the highest rates of COVID-19 transmission combined with positive acceleration, jerk, and 7-day persistence. Wisconsin and Illinois also have alarming indicators and already lead the nation in daily new COVID-19 infections. As the United States enters its third wave of COVID-19, all 50 states and the District of Colombia have positive rates of speed between 7.58 (Hawaii) and 175.01 (North Dakota), and persistence, ranging from 4.44 (Vermont) to 195.35 (North Dakota) new infections per 100,000 people.

Conclusions: Standard surveillance techniques such as daily and cumulative infections and deaths are helpful but only provide a static view of what has already occurred in the pandemic and are less helpful in prevention. Public health policy that is informed by dynamic surveillance can shift the country from reacting to COVID-19 transmissions to being proactive and taking corrective action when indicators of speed, acceleration, jerk, and persistence remain positive week over week. Implicit within our dynamic surveillance is an early warning system that indicates when there is problematic growth in COVID-19 transmissions as well as signals when growth will become explosive without action. A public health approach that focuses on prevention can prevent major outbreaks in addition to endorsing effective public health policies. Moreover, subnational analyses on the dynamics of the pandemic allow us to zero in on where transmissions are increasing, meaning corrective action can be applied with precision in problematic areas. Dynamic public health surveillance can inform specific geographies where quarantines are necessary while preserving the economy in other US areas.

(J Med Internet Res 2020;22(12):e24286) doi:10.2196/24286

KEYWORDS

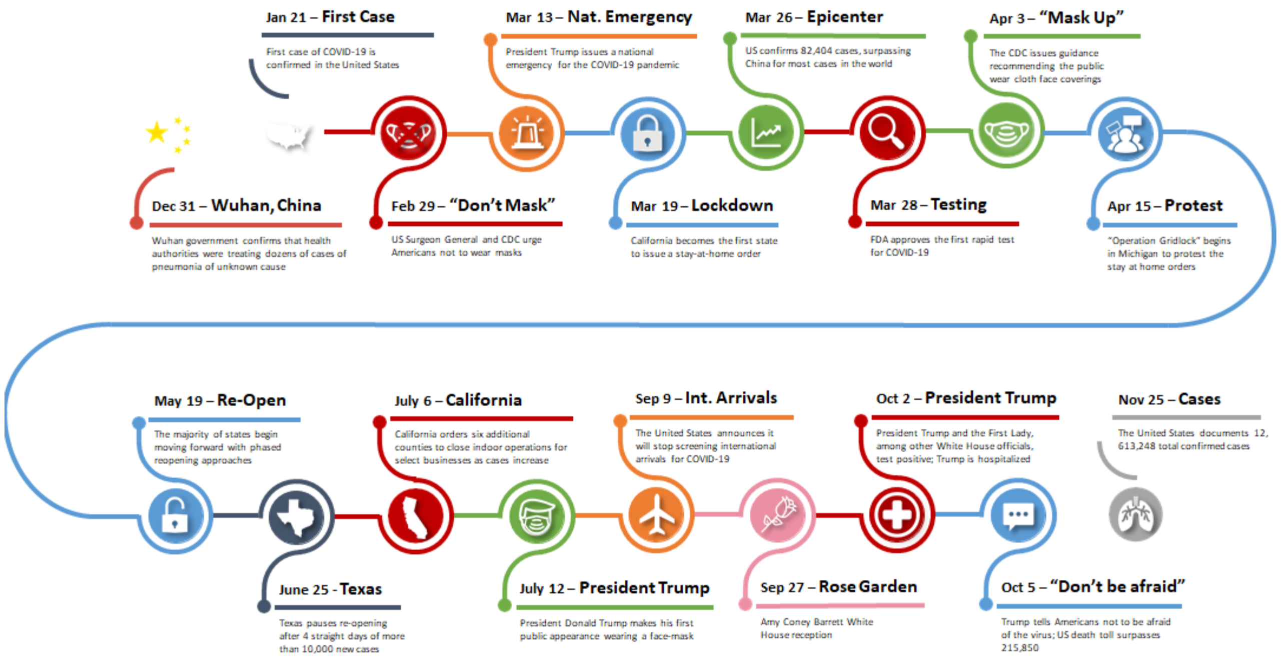
global COVID-19 surveillance; United States public health surveillance; US COVID-19; surveillance metrics; dynamic panel data; generalized method of the moments; United States econometrics; US SARS-CoV-2; US COVID-19 surveillance system; US COVID-19 transmission speed; COVID-19 transmission acceleration; COVID-19 speed; COVID-19 acceleration; COVID-19 jerk; COVID-19 persistence; Arellano-Bond estimator; COVID-19

Introduction

Background

The emergence of SARS-CoV-2 in 2019 led to its associated disease, COVID-19, becoming one of the most severe and widespread disease outbreaks in modern history [1]. On January 21, 2020, the first confirmed case of COVID-19 was recorded in the United States (see Figure 1) [2]. Over the next 2 months, cases spread worldwide at an alarming rate, leading the World Health Organization on March 11, 2020, to officially recognize COVID-19 as a pandemic [3]. Around the world, countries quickly implemented public health policies to mitigate the public health and economic impacts caused by the virus. Nations have had varying levels of success at controlling the transmission of COVID-19 due to numerous political, cultural, economic, and structural factors [4]. By November 13, 2020, the daily number of new COVID-19 cases in the United States reached 177,224, the highest number for any single day recorded since the pandemic began, contributing to the 11,485,176 cumulative cases of people infected by COVID-19 and the 250,029 people who have died from it [5].

Figure 1. Timeline of COVID-19 in the United States. CDC: Centers for Disease Control and Prevention; FDA: Food and Drug Administration.



It is logical that the United States experienced technical difficulties at the beginning of the pandemic because we were dealing with a new virus, a pandemic unprecedented in modern times, and uncharted territory [6], which explains why the United States became the third global epicenter of the pandemic due to a delayed response [7]. It is illogical that the United States continues to underperform every country worldwide in controlling COVID-19 because of our wealth of resources and that we now know the epidemiological risk factors, cultural norms, and politicized public health policies that result in case clustering in the United States [4]. Unfortunately, the current US response to COVID-19 parallels President Wilson's response to the 1918 Influenza Pandemic that killed 675 million Americans because of no national response to control the outbreak, misinformation, unclear communication with the public, and widespread mistrust and panic [8-13].

Public Health Policies and Governance

Without a national plan to control COVID-19, some mayors and governors implemented policies such as public masking mandates, prohibitions on gatherings, school closings, restrictions on commercial activities, and broad social distancing measures in an attempt to "flatten the curve" [1,14,15]. Empirical evidence supports social distancing, quarantines, and wearing masks as a means to stop the transmission of COVID-19 [14,16-18]. Leaders who minimize the pandemic, such as White House Press Secretary Kayleigh McEnany, who stated that rising cases are nothing more than "embers that need to be put out" [19], have fostered a false optimism bias that undermines public health guidelines [20]. Framing adherence to those guidelines as a civil liberties issue discourages safe practices, such as advisories by Trump coronavirus advisor Scott Atlas, who urged Michigan to *rise up* against new COVID-19 policies put in place by Governor Whitmer [21]. Conflicting messaging and politicization of COVID-19 have fueled the pandemic. Although the United States only comprises 4% of the global population, it exceeds 20% of the global COVID-19 caseload [22]. Presently, the world is implementing policies to control the second wave of COVID-19, while the United States is responding to its third wave [23-25].

Research has explained why the Unites States consistently leads the world in COVID-19 cases and how we lost control of the outbreak; however, what has not been widely recognized is the need for ongoing systematic public health surveillance to gain control of the pandemic [26-28]. Surveillance informs prevention efforts and signals whether policies put in place to combat COVID-19 are effective or where outbreaks are occurring unabated [26,27]. There are several excellent sources of surveillance data that enumerate the daily and cumulative number of new infections and deaths, the rates of infections and deaths, as well as moving 7-day averages of infections and deaths [29]. Although these indicators are useful, they also have incomplete case ascertainment [26-28]. Without universal testing, surveillance systems are limited to only those cases that are reported from those who are tested, hence erring toward individuals with more severe symptoms [26-28]. Moreover, traditional surveillance metrics are static and provide a proxy of the pandemic for what already has occurred while omitting the pandemics' dynamics that will inform future effective public health guidelines at the epicenter of outbreaks [30]. Dynamic surveillance allows public health leaders to be proactive rather than reactive, which can prevent novel infections. Without universal testing and a national plan to control the pandemic, public health surveillance is even more essential. Moreover, those leaders concerned with choosing between protecting the public from COVID-19 versus protecting the economy require dynamic public health surveillance to understand which parts of each state can remain open and which must go into quarantine [31-34].

Objective

The objective of this paper is to use static and dynamic surveillance metrics to measure the caseload and dynamics of the COVID-19 pandemic at the national and state level [30,35-38]. The basic question we are trying to inform is: how are we doing this week relative to previous weeks? From a public health perspective, we need to understand if there are more cases per day this week than last week, the number of new cases is increasing from day to day, and the number of day-to-day increases in the number of cases is bigger this week than last week. Additionally, we would like some indicative information about significant shifts in how the pandemic is progressing—positive shifts could be the first indicators of the emergence of a new or recurrent hot spot, and negative shifts could be the first indicators of a successful public health policy.

Methods

This study relies on a longitudinal trend analysis study design to investigate the transmission of COVID-19 at the national and the state level over time. The COVID Tracking Project [39] compiles data found on the web [40] from public health departments, universities, and the media to track COVID-19 [3,29,41-43]. Data for the past several months were accessed from the GitHub repository [44]. This resulted in a balanced panel of all 50 states and the District of Columbia (51 x 46 days = 2346). There were in total 3927 observations used to estimate 46 days of data for US states in this analysis. An empirical difference equation was specified in which the number of positive cases in each state at each day is a function of the prior number of cases, the level of testing, and weekly shift variables that measure whether the contagion was growing faster, the same, or slower than the previous weeks. This resulted in a dynamic panel model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in R (The R Foundation for Statistical Computing) [45,46].

Arellano-Bond estimation of difference equations has several statistical advantages: it allows for statistical examination of the model's predictive ability and the validity of the model specification, it corrects for autocorrelation and heteroscedasticity, it has good properties for data with a small number of time periods and large number of states, and it corrects for omitted variables issues and provides a statistical test of correction validity. With these advantages, the method is applicable to ascertaining and statistically validating changes in the evolution of the pandemic within a period of a week or less, such as changes in the reproduction rate. Oehmke et al [28,30] provide a detailed discussion of the methods. Finally, we calculated these indicators to inform public health leaders where to take corrective action at a local level.

To ascertain whether pandemic growth was explosive, we examined whether the acceleration and jerk were positive and if the persistence effect was positive as well as if it was larger than the speed. We examined these indicators week over week for 7 weeks along with the daily caseload. The persistence effect is an indicator for mathematically explosive growth (ie, it indicates if the difference equation has a solution that lies outside the unit circle), but large positive acceleration and jerk as well as persistence are indicative of explosive growth in a practical sense (ie, the COVID-19 caseload is expected to increase much more rapidly than in the current or recent week). However, this is a forward-looking statement and must be interpreted with due circumspection. To discuss potential explosive growth, we looked at five data points including speed, acceleration, jerk, persistence, and daily caseload each week for 7 weeks.

Results

State Regression Results

We present regression results for the panel comprising all 50 US states and the District of Columbia in Table 1. The subsequent weekly surveillance products were based on these regressions.

Table 1. Arellano-Bond dynamic panel data model of COVID-19 dynamics at the state level.

Variable	Values	*P* value
7-day lag, coefficient	1.09	<.001
Cumulative tests, coefficient	0.01	.48
7-day lag shift, coefficient	0.108	<.03
Wald statistic for regression, chi-square (*df*)	13,504 (8)	<.001
Sargan statistic for validity, chi-square (*df*)	48.3 (367)	>.99

The regression Wald statistic was statistically significant (χ^2_8=13,504, *P*<.001). The Sargan test was insignificant, failing to reject the validity of overidentifying restrictions (χ^2_{367}=48.3, *P*>.99).

The coefficients on the 7-day lag were both positive and statistically significant (*P*<.001), demonstrating that the number of infections 1 week prior to data collection had a significant effect on number of infections at the point of data collection. Cumulative tests also had a slightly positive but insignificant effect (coefficient 0.01, *P*=.48). The shift parameter 14 days ago was positive and significant (coefficient 0.108, *P*<.03), suggesting that exogenous shift events had a positive effect on cases.

Interpretation: Regression Results

The lagging indicators and shift parameters suggested recent change in disease transmission in the United States between September 28 and November 15, 2020. Specifically, the most recent 7-day lag this week was 10% faster than last week. The shift in the most recent 14 days, or 2 weeks, was positive and significant (*P*<.03).

Surveillance Results

Table 2 demonstrates the dynamics of the pandemic for the leading six states in speed, acceleration, jerk, 7-day persistence, and 7-day moving average over the course of the 7-week study beginning on September 28 and ending November 15, 2020. Table 2 is truncated and only lists the leading six states for our metrics. For a complete view of each state, Multimedia Appendix 1 contains Tables S5-S18 that list the traditional surveillance metrics as well as the dynamic surveillance metrics of speed, acceleration, and 7-day persistence rate.

Table 2. State pandemic dynamics

Date, State	Speed	State	Acceleration	State	Jerk	State	7-day persistence effect	State	7-day moving average
Oct 1, 2020									
ND	51.9	SD	4.6	SD	5.4	ND	55.7	TX	4083.1
SD	49.1	AK	2.6	AK	1.6	SD	42.4	CA	3292.4
WI	42.7	ID	1.7	MO	1.5	WI	37.7	WI	2489.0
UT	30.9	MT	1.3	ID	1.2	UT	31.4	FL	2250.7
MT	30.2	WI	1.3	TN	1.0	OK	31.2	NC	2102.9
AR	27.3	MO	1.0	NV	0.8	AR	30.1	IL	2051.4
Oct 8, 2020									
ND	59.4	ND	2.9	VA	2.7	ND	56.5	TX	4184.7
SD	52.9	MT	2.5	ND	2.0	SD	53.4	CA	3016.1
WI	42.6	VA	2.3	IA	1.8	WI	46.6	WI	2477.7
MT	42.5	UT	2.2	UT	1.7	UT	33.6	FL	2363.7
UT	35.2	SC	1.9	RI	1.6	MT	32.9	IL	2215.7
ID	29.6	TN	1.5	NJ	1.4	AR	29.8	NC	1784.0
Oct 15, 2020									
ND	80.0	SD	4.7	SD	7.2	ND	64.6	TX	4002.0
SD	74.3	MS	3.6	MS	6.7	SD	57.6	CA	3371.6
MT	56.9	ND	3.4	LA	4.6	WI	46.3	WI	3093.6
WI	53.1	NE	2.1	KY	4.4	MT	46.2	IL	3031.3
UT	37.9	WI	1.9	AK	3.1	UT	38.3	FL	2648.1
NE	37.4	AL	1.8	MT	3.0	ID	32.3	NC	1934.6
Oct 22, 2020									
ND	101.2	AL	3.9	SD	7.5	ND	87.1	TX	5041.7
SD	80.8	ND	3.4	MO	7.4	SD	80.9	IL	4155.4
MT	62.9	SD	2.8	ND	7.2	MT	61.9	WI	3528.7
WI	60.6	MT	2.7	AL	6.2	WI	57.9	FL	3231.6
ID	46.0	ID	2.7	MT	2.5	UT	41.3	CA	3189.0
NE	43.1	RI	2.7	FL	1.9	NE	40.8	TN	2154.3
Oct 29, 2020									
ND	113.9	ND	6.3	WI	4.3	ND	110.2	TX	5960.0
SD	112.8	WI	3.7	CT	3.0	SD	88.0	IL	5203.6
WI	75.8	MN	3.3	WY	1.8	MT	68.5	CA	4597.1
MT	69.7	CT	3.3	NV	1.3	WI	66.0	WI	4411.7
WY	58.9	MO	2.8	MO	1.2	ID	50.1	FL	3717.9
AK	51.7	MI	2.7	MN	1.2	NE	47.0	MI	2852.4
Nov 5, 2020									
ND	163.2	IA	9.3	SD	11.2	ND	134.6	IL	7654.1
SD	131.6	ND	5.9	IA	7.0	SD	132.9	TX	6882.0
WI	90.5	SD	5.8	MS	4.2	WI	89.5	WI	5266.6
MT	81.7	NE	4.9	RI	2.4	MT	82.1	FL	4577.9
WY	71.6	UT	4.3	IL	2.4	WY	69.4	CA	4524.6
IA	71.3	IL	4.0	KY	2.2	AK	61.2	MI	4031.4

Date, State	Speed	State	Acceleration	State	Jerk	State	7-day persistence effect	State	7-day moving average
Nov 12, 2020									
ND	175.0	WY	18.3	WY	25.5	ND	195.3	IL	11,827.4
SD	154.5	SD	10.7	LA	12.8	SD	157.5	TX	8406.7
WY	125.1	LA	9.7	ND	6.5	WI	108.3	CA	6719.0
WI	112.9	MN	8.3	MN	5.6	MT	97.8	WI	6571.6
IA	112.6	UT	5.0	UT	4.0	WY	85.7	MI	5845.7
NE	103.8	ND	4.9	SD	3.8	IA	85.3	OH	5612.4

Although Texas and California ranked in the top six states for daily average of new infections for 7 consecutive weeks, they did not rank high for rates of speed, acceleration, jerk, or 7-day persistence. Florida ranked in the top for daily average of new infections for 6 of the 7 weeks, and in the middle of the study during the week of October 22, 2020, Florida had a positive jerk of 1.9 per 100,000 people, indicating that the pandemic was not just accelerating but accelerating at an increasing rate. North Carolina ranked in the top six states for novel infections for 3 of the 6 weeks but did not rank high for rates of speed, acceleration, jerk, or persistence. Illinois was ranked in the top six states for novel infections all 7 weeks; during the week of November 5, Illinois ranked high in acceleration at 4.0 additional new daily infections per 100,000 people per day, or 28 per week, and had a positive jerk of 2.4 infections per 100,000 people during the same week, indicating that acceleration increased from 1.6 in the prior week to 4.0 the week of November 5.

In Table 2, we examine speed, acceleration, jerk, persistence, and caseload each week for 7 weeks for the top six states for each metric. Dynamic metrics for North Dakota during the 7-week study were alarming with 22 positive data points that indicated explosive growth even though North Dakota never made the top six list of largest case load. North Dakota had the highest rate in the United States in terms of speed of new infections all 7 weeks. The pandemic accelerated 6 of the 7 weeks. The jerk was positive 3 of the 7 weeks. North Dakota had the highest persistence rate all 7 weeks, meaning new cases today were statistically attributable to novel infections a week ago. There was some underlying condition or events that have echoed forward with increasing numbers of new infections each week. During the week of November 12, 2020, North Dakota had the highest speed of infections in the United States at 175 new infections per 100,000 people; measures of acceleration indicated the speed of infections was increasing at 4.9 new daily infections per 100,000 people per day, which over a week cumulates to over 34 more new daily infections per 100,000 people than a week ago. The jerk of 6.5 new infections per 100,000 people indicates that the state moved from a deceleration the prior week to a rapid acceleration the week of November 12. Finally, North Dakota had the highest persistence effect in the United States at an average of 195.3 new infections per day for the week of November 12 that were statistically attributable to infections the prior week. Not only was North Dakota experiencing explosive growth, the surveillance indicators suggested that the outbreak is still strengthening. Corrective action is needed immediately.

South Dakota has 24 positive upward trending data points for speed, acceleration, jerk, and persistence week over week for 7 weeks, indicating explosive growth. South Dakota had the second highest rate of novel infections and persistence in the United States; for 5 of the 7 weeks, North and South Dakota's acceleration and jerk were positive, indicating that the pandemic was strengthening at an increasing rate. Utah has 12 data points in Table 2 of the leading six states in speed, acceleration, jerk, and persistence. During the past week of November 12, 2020, Utah had positive acceleration and jerk. Montana had 17 data points, indicating explosive growth. Montana was a top-six state in speed for 6 of the last 7 weeks, acceleration 3 of the last 7 weeks, and jerk 2 of the last 7 weeks. Utah ranked in the top six states for persistence for 6 of the last 7 weeks.

Arizona is more promising. Although Arizona ranked sixth in the rate of new infections during the week of October 1, 2020, and had a leading persistence rate for the weeks of October 1 and October 8, Arizona does not appear in the leading states in Table 2 afterward. Iowa had six indicators and Idaho and Nebraska each had seven indicators of explosive growth.

Wisconsin had a high caseload, and it was in the top six states for speed, acceleration, and persistence. Wisconsin remained a leader in speed across all 7 weeks and cumulatively had the highest number of new infections per 100,000 people over those weeks. Wisconsin's pandemic accelerated in 3 of the 7 weeks. During the week of October 1, 2020, Wisconsin had a speed of 42.7, and by the week of November 12, 2020, its speed had reached 112.9. Wisconsin ranked as a leading state all 7 weeks for persistence, indicating there was some underlying condition that persisted and echoed forward.

Wyoming showed the largest acceleration over the 7 weeks, with speed accelerating from 19.2 new daily cases per 100,000 people during the week of October 1, 2020, to 125.1 during the week of November 12, 2020.

Table 2 demonstrates the utility of dynamic surveillance metrics because traditional surveillance metrics only report caseloads of each state. Larger states (Table 3) will more likely be captured in those types of metrics, missing the dynamics of the pandemic and missing those states who demonstrate high rates of infection speed, speeds that are accelerating, accelerating speeds that jerk upwards, and those new cases today that are statistically attributable to new cases last week or those new cases with an underlying biological or social condition (eg, attending mass events without face masks or social distancing), that is, echoing forward and leading to additional infections.

Table 3. Most populous US states.

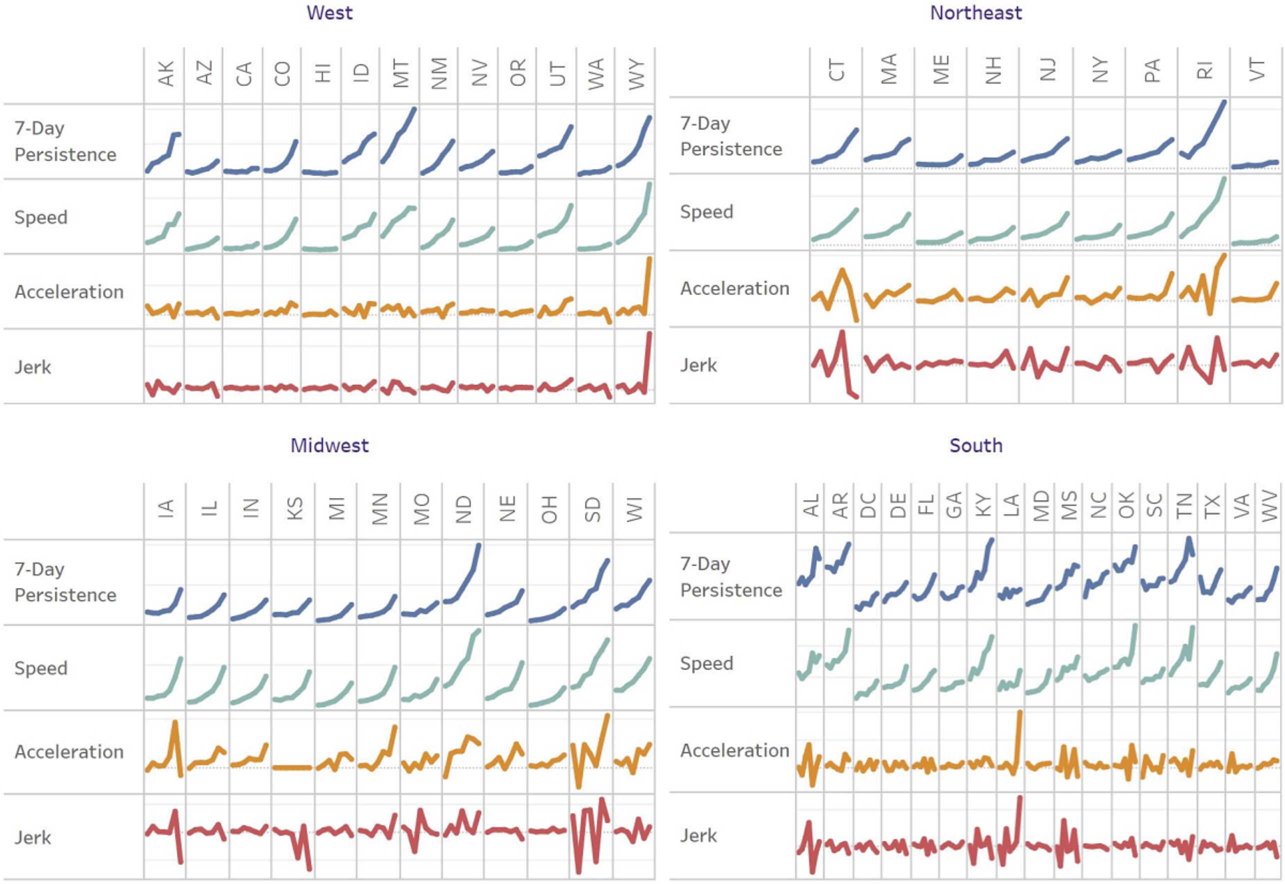
State	Population as of 2020
California	39,937,500
Texas	29,472,300
Florida	21,993,000
New York	19,440,500
Pennsylvania	12,820,900
Illinois	12,700,381

Table 4 provides the average of the US states or the US rates for speed, jerk, and 7-day persistence. During the week of October 1, 2020, the speed was positive at 12.91 infections per 100,000 people. In the speed column, for each subsequent week in the study, the speed increased to a maximum speed of 39.38 during the week of November 11, 2020. This explains why the acceleration rate per 100,000 people increased weekly, as acceleration was the measure of increases in speed week over week. In the final column, we see the 7-day persistence begins at 13.99 per 100,000 people during the week of October 1 that are statistically attributed to the number of novel infections 7 days earlier. By the week of November 12, this persistence had increased to 33.62, meaning that 85% of the 39.38 average daily new cases the week of November 12 were statistically attributable to new cases in the prior week. This could happen if, for example, in-restaurant diners last Saturday became infected and then dined in-restaurant again this Saturday and infected other diners, who then eat in-restaurant next Saturday and infect the next cohort, etc. Sequential sports, election, and postelection watch parties could generate the same effect. The novel measures, thus, suggest that a return to stricter guidelines is appropriate for regaining control of the pandemic and were indicated as early as the week of October 22, 2020, which was the third consecutive week of positive acceleration and jerk, and increasing persistence. Weekly trends by state are displayed in Figure 2. See Multimedia Appendices 2-5 for US maps displaying weekly acceleration, jerk, and persistence.

Table 4. US pandemic dynamics.

Date	Speed	Acceleration	Jerk	7-day persistence effect
Oct 1, 2020	12.91	0.12	–0.12	13.99
Oct 8, 2020	13.85	0.41	0.13	14.05
Oct 15, 2020	16.03	0.34	0.11	15.08
Oct 22, 2020	18.39	0.45	0.24	17.45
Oct 29, 2020	22.97	0.63	–0.16	20.03
Nov 5, 2020	28.09	1.24	0.18	27.11
Nov 12, 2020	39.38	1.45	–0.34	33.62

Figure 2. Weekly US state statistics.



Discussion

Principal Results

The United States has had an uncoordinated, decentralized response to COVID-19. Although traditional public health surveillance provides a static view of the pandemic, the data are limited by secondary bias from undercounting, reporting delays, testing issues, and other forms of contamination. The novel measures presented in this study take steps toward resolving these shortcomings and measure the dynamics of the pandemic. They also provide greater insight into the evolution of a pandemic, such as where COVID-19 is transmitted and whether rates of transmission are increasing. Measures like 7-day persistence control for incomplete case ascertainment and look retrospectively to understand current infection rates, and speed, acceleration, and jerk provide a dynamic perspective on future cases.

The data presented in Tables S5-S18 in Multimedia Appendix 1 indicate that some states are responding well to the pandemic, while others are responding poorly in the present or suffering from past mistakes. States can be examined systematically to determine their relative performance. At a high level, the static metrics demonstrate national and state burden of disease over a 7-week period. Acceleration and jerk indicate the potential progress of disease burden. A state with a high number of new cases but a negative acceleration and jerk is more likely to have a good trajectory than a state with a high number of cases and a positive acceleration and jerk. The 7-day persistence metric hints at both past decision making and future progress. A state with a high 7-day persistence may have experienced a super-spreader, causing an echo in transmissions to the present and potentially into the future. These echoes may be particularly susceptible to control by identifying the causal event and performing rigorous contact tracing or eliminating the offending behaviors (eg, gatherings without face masks and social distancing).

Overall, the state of the pandemic in the United States is concerning based on static surveillance and dynamic metrics. All 50 states and the District of Colombia have positive rates of new infections. Moreover, the rates of infection are higher and faster than at any other time since the onset of the pandemic.

Why the COVID-19 Response Differs From Previous Infectious Disease Outbreaks?

COVID-19 has been politicized and polarized, flamed by social media and misinformation [47-56]. Scientists have consistently emphasized that a reduction in COVID-19 transmission is predicated on social distancing, quarantines, hand hygiene, crowd control, and wearing masks [14,16,17]. Others weighed COVID-19 safety measures against shutting down the economy to prevent financial hardships that dynamic surveillance can inform. In the absence of policy actions, COVID-19 experienced exponential growth rates of approximately 38% per day [57]. Countries with anticontagion policies significantly and substantially slowed the growth of COVID-19 [57], yet many US states are unable to implement COVID-19 control policies. Finally, since the start of the COVID-19 pandemic, there was confusion about who is “in charge” of the US epidemic response

[58]. Federal government powers do not allow it to impose specific restrictions on states [59]; however, its resources are *vast*, with the ability to develop national guidelines, promote scientific research and vaccine development, and direct resources toward COVID-19 relief [60-62]. Unfortunately, city mayors and state governors must patch together uncoordinated piecemeal COVID-19 control policies [63-65].

Conflicts in Adopting Policy

The US COVID-19 epidemic has rebounded and is accelerating rapidly with multiple outbreaks in most US states, indicating shortfalls in preparedness and a dearth of nonpharmaceutical control measures. Although standard surveillance techniques such as daily and cumulative infections and deaths are helpful, they have incomplete case ascertainment, err on the side of the most severe cases, and provide a static view of what has already occurred during the pandemic, which is less helpful in prevention. Public health policy that is informed by dynamic surveillance can shift the country from reacting to COVID-19 transmissions to being proactive and taking corrective action when indicators of speed, acceleration, jerk, and persistence remain positive week over week. Implicit within our surveillance is an early warning system that indicates when there is problematic growth in COVID-19 transmissions as well as to signal when growth will become explosive without action.

A public health approach that focuses on prevention can prevent major outbreaks in addition to endorsing effective public health policies and may resolve conflict between various government entities. Moreover, subnational analyses on the dynamics of the pandemic allows us to zero in on where transmissions are increasing, meaning corrective action can be applied with precision in problematic areas. Dynamic public health surveillance can inform specific geographies where quarantines are necessary, preserving the economy in other US areas.

Without a unified national policy to address COVID-19, individual cities and states have launched efforts to control the spread of the pandemic. Similar studies have found that states that imposed strict guidelines saw drastic reductions in COVID-19 spread and avoided significant increased case counts and fatalities [66,67].

Discord between New York City Mayor Bill De Blasio and New York Governor Andrew Cuomo has resulted in delays in shutting down new virus hot spots in neighborhoods across New York City [68]. In other states, governors have faced problems passing public health safety measures with proposed legislation ruled unconstitutional by state courts. In early May, the Wisconsin Supreme Court overturned a statewide stay-at-home order put in place by Governor Evers to control rising cases [69]. Wisconsin has since faced a large increase in COVID-19 cases and reports some of the largest numbers of new infections per day. Recently, Governor Whitmer of Michigan also experienced issues with an executive order she passed when the Michigan Supreme Court ruled her statewide mask mandate as unconstitutional [70]. Even within states, some cities have imposed mask mandates on their own accord, while others have refused, leading to different outcomes. In South Carolina, areas with mask mandates reported a 46% greater reduction in COVID-19 case rates than those areas without mask mandates [71]. The Tri-State area of New Jersey, New York, and Connecticut coordinated travel restrictions, public health guidelines, and economic activities [72]. After initial success, four additional states joined to quell the COVID-19 pandemic [73]. These coordination challenges and frictions between levels of government have been observed in other countries as well, including Brazil, India, and Nigeria [74], and can be especially pronounced in settings of vertically divided authority where subnational governments are led by opposition parties. Germany is a notable exception; states have control over their own health systems and each state was able to pursue its own testing without waiting for approval from a national health laboratory—a factor that has been partially credited for the country's swift testing response [75].

Current State of the Pandemic

An important change in the dynamic of COVID-19 transmission occurred starting mid-August when universities and schools around the country started to open for their new academic year [30]. Approximately 53 million US children between the ages of 5-17 years resumed school in fall of 2020 [76]. The CDC confirmed over 2800 COVID-19 cases at the University of Wisconsin within the first month of reopening [77]. Similarly, the University of Missouri has confirmed over 1600 COVID-19 cases among its students since reopening on August 19, 2020 [78]. Reinstating college sports also resulted in novel infections [79].

States such as Wisconsin and Indiana have reported pandemic highs in daily COVID-19 new infection counts [3,80]. As October 2020 began, Utah reported the nation's highest positivity rate [81]. This trend was high throughout September, suggesting a potential dearth in testing that caused an undercount [81]. Florida, which saw a surge increase over the summer, has seen a steady decline in its daily case counts after stricter measures were put in place by the governor; however, these measures have since been lifted [82]. Wisconsin, Iowa, and Utah have not instituted mask mandates and have seen worsening infection counts that have not improved. The four states of greatest concern are Wisconsin, North Dakota, South Dakota, and Illinois. Wisconsin, North Dakota, and South Dakota have multiple data points indicating explosive growth. Wisconsin and Illinois both top the nation in daily new infections and have positive growth rates. California and Pennsylvania have positive speed, acceleration, jerk, and persistence, suggesting the outbreak in these two populous states will be significant in terms of caseload.

Conclusion

The variation in speed, acceleration or deceleration, and jerk between and within states is consistent with varying degrees of compliance with public health guidelines to combat COVID-19 in terms of social distancing, masks, hand hygiene, and crowd control. Some states have overall higher caseloads because they have higher populations, such as California, New York, Illinois, Texas, and Florida, and others have alarming speed, acceleration, and jerk. Nationally, the US COVID-19 epidemic has rebounded and is accelerating rapidly in multiple states, indicating shortfalls in preparedness and a dearth of nonpharmaceutical control measures. Although standard surveillance techniques such as daily and cumulative infections and deaths are helpful, they also have incomplete case ascertainment, err on the side of the most severe cases, and provide a static view of what has already occurred in the pandemic, which is less helpful for future planning. Public health policy that is informed by dynamic surveillance can shift the country from reacting to COVID-19 infections and deaths to being proactive and taking corrective action when indicators of speed, acceleration, jerk, and persistence remain positive week over week.

Implicit within our surveillance is an early warning system that indicates when there is problematic growth in COVID-19 transmissions as well as signals when growth is likely. For example, at the US level, the novel metrics indicated a need for strengthening public health measures as early as the week of October 22, 2020, when the highest numbers of new daily cases were 82,000-84,000 per day. In the absence of these novel metrics, little action was taken until the number of new cases had more than doubled to over 161,934 new cases on November 17, 2020 [5]. At the state level, the indicators suggested that explosive growth occurring in the Dakotas, Wisconsin, and Wyoming could also have been identified in late October. Public health actions taken at that time in those and other states and areas might have significantly reduced the number of new cases and prevented the severe overtaxing of hospital and medical resources that is now happening.

A public health approach that focuses on prevention can prevent major outbreaks in addition to confirming when public health guidelines are effective and controlling the pandemic. Moreover, subnational analyses on the dynamics of the pandemic allows us to zero in on where transmissions are increasing, meaning corrective action can be applied with precision on problematic areas. However, this approach requires subnational, dynamic public health surveillance that can inform specific geographies where lockdowns or other measures are necessary. This paper provides novel surveillance measures that can help fill that exact need.

Limitations

Our data are limited by state-level granularity and differences in testing and reporting within states. State-level granularity, although superior to national reporting, provides less detail than county- or city-level reporting. This limitation is particularly pronounced in states with large urban centers governed by powerful mayors. Testing and reporting varies across and within states for many reasons, including the decentralization of US health care. Both insurers and providers have inconsistent policies in place and resources deployed for COVID-19 testing. To address this need for small area surveillance, we have generated static and dynamic surveillance for larger metropolitan areas in the United States in an additional publication.

Comparison With Prior Work

This study is part of a broader research program at Northwestern Feinberg School of Medicine, The Global SARS-CoV-2 Surveillance Project: Policy, Persistence, & Transmission. This research program developed novel surveillance metrics to include rates of speed, acceleration, jerk, and 7-day persistence [28,30]. We have also derived surveillance metrics for all global regions.

Acknowledgments

This study and the resulting surveillance system were funded in part by the Davee Innovations Research Endowment for North America. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors alone and do not necessarily reflect the opinions of funders or authors' employers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Tables S5-S18.
[DOCX File , 69 KB - jmir_v22i12e24286_app1.docx]

Multimedia Appendix 2
Explosive growth potential.
[PNG File , 460 KB - jmir_v22i12e24286_app2.png]

Multimedia Appendix 3
US weekly SARS-CoV-2 trends.
[PNG File , 263 KB - jmir_v22i12e24286_app3.png]

Multimedia Appendix 4
Weekly US 7-day persistence map.
[PNG File , 496 KB - jmir_v22i12e24286_app4.png]

Multimedia Appendix 5
Weekly US acceleration and jerk.
[PNG File , 474 KB - jmir_v22i12e24286_app5.png]

References

1. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
2. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Washington State 2019-nCoV Case Investigation Team. First case of 2019 novel coronavirus in the United States. N Engl J Med 2020 Mar 05;382(10):929-936 [FREE Full text] [doi: 10.1056/NEJMoa2001191] [Medline: 32004427]
3. Mahase E. Covid-19: WHO declares pandemic because of "alarming levels" of spread, severity, and inaction. BMJ 2020 Mar 12;368:m1036. [doi: 10.1136/bmj.m1036] [Medline: 32165426]
4. Chin T, Kahn R, Li R, Chen JT, Krieger N, Buckee CO, et al. US-county level variation in intersecting individual, household and community characteristics relevant to COVID-19 and planning an equitable response: a cross-sectional analysis. BMJ Open 2020 Sep 01;10(9):e039886 [FREE Full text] [doi: 10.1136/bmjopen-2020-039886] [Medline: 32873684]
5. COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-11-25]
6. Patel R, Babady E, Theel E, Storch GA, Pinsky BA, St George K, et al. Report from the American Society for Microbiology COVID-19 International Summit, 23 March 2020: value of diagnostic testing for SARS-CoV-2/COVID-19. mBio 2020 Mar 26;11(2):e00722-e00720 [FREE Full text] [doi: 10.1128/mBio.00722-20] [Medline: 32217609]
7. Nathavitharana RR, Patel PK, Tierney DB, Mehrotra P, Lederer PA, Davis S, et al. Innovation and knowledge sharing can transform COVID-19 infection prevention response. J Hosp Med 2020 May;15(5):299-301. [doi: 10.12788/jhm.3439] [Medline: 32379035]
8. Kolata G. Flu: The Story Of The Great Influenza Pandemic of 1918 and the Search for the Virus that Caused It. New York, NY: Atria Books; Jan 09, 2001.
9. Flecknoe D, Charles Wakefield B, Simmons A. Med Confl Surviv 2018 Jun;34(2):61-68. [doi: 10.1080/13623699.2018.1472892] [Medline: 29764189]
10. Workshop proceedings pandemic influenza? Past, present, future: communicating today based on the lessons from the 1918-1919. Centers for Disease Control and Prevention. 2006 Oct 17. URL: https://www.cdc.gov/flu/pandemic-resources/pdf/workshop.pdf [accessed 2020-10-01]
11. Davis K. More Deadly Than War: The Hidden History of the Spanish Flu and the First World War. New York, NY: Henry Holt and Co; May 15, 2018.
12. Radusin M. The Spanish flu--part II: the second and third wave. Vojnosanit Pregl 2012 Oct;69(10):917-927. [Medline: 23155616]
13. Spinney L. Pale Rider: The Spanish Flu of 1918 and How It Changed the World. New York, NY: PublicAffairs; Sep 12, 2017.
14. Laestadius L, Wang Y, Ben Taleb Z, Kalan ME, Cho Y, Manganello J. Online national health agency mask guidance for the public in light of COVID-19: content analysis. JMIR Public Health Surveill 2020 May 26;6(2):e19501 [FREE Full text] [doi: 10.2196/19501] [Medline: 32427103]
15. Berman E. The roles of the state and federal government in a pandemic. SSRN J Preprint posted online June 2, 2020. [doi: 10.2139/ssrn.3617058]
16. Ratzan S, Sommariva S, Rauh L. Enhancing global health communication during a crisis: lessons from the COVID-19 pandemic. Public Health Res Pract 2020 Jun 30;30(2):3022010. [doi: 10.17061/phrp3022010] [Medline: 32601655]
17. Esposito S, Principi N, Leung CC, Migliori GB. Universal use of face masks for success against COVID-19: evidence and implications for prevention policies. Eur Respir J 2020 Jun;55(6):2001260 [FREE Full text] [doi: 10.1183/13993003.01260-2020] [Medline: 32350103]
18. Considerations for implementing and adjusting public health and social measures in the context of COVID-19. World Health Organization. 2020 Nov 04. URL: https://www.who.int/publications/i/item/considerations-in-adjusting-public-health-and-social-measures-in-the-context-of-covid-19-interim-guidance [accessed 2020-11-15]
19. Press briefing by Press Secretary Kayleigh McEnany. The White House. 2020 Nov 20. URL: https://www.whitehouse.gov/briefings-statements/press-briefing-press-secretary-kayleigh-mcenany-11-20-2020/ [accessed 2020-11-21]
20. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
21. LeBlanc P, Diamond J. Trump coronavirus adviser Scott Atlas urges Michigan to 'rise up' against new Covid-19 measures. CNNPolitics. 2020 Nov 15. URL: https://www.cnn.com/2020/11/15/politics/scott-atlas-coronavirus-michigan/index.html [accessed 2020-11-15]

22. COVID-19 coronavirus pandemic. Worldometer. 2020. URL: https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?%22 [accessed 2020-08-15]
23. Solis J, Franco-Paredes C, Henao-Martínez AF, Krsak M, Zimmer S. Structural vulnerability in the U.S. revealed in three waves of COVID-19. Am J Trop Med Hyg 2020 Jul;103(1):25-27 [FREE Full text] [doi: 10.4269/ajtmh.20-0391] [Medline: 32383432]
24. Seligmann H, Vuillerme N, Demongeot J. Summer COVID-19 third wave: faster high altitude spread suggests high UV adaptation. medRxiv Preprint posted online August 21, 2020. [doi: 10.1101/2020.08.17.20176628]
25. Cacciapaglia G, Cot C, Sannino F. Second wave COVID-19 pandemics in Europe: a temporal playbook. Sci Rep 2020 Sep 23;10(1):15514. [doi: 10.1038/s41598-020-72611-5] [Medline: 32968181]
26. Thacker S, Berkelman R. Public health surveillance in the United States. Epidemiol Rev 1988;10:164-190. [doi: 10.1093/oxfordjournals.epirev.a036021] [Medline: 3066626]
27. Lee LM, Thacker SB, St. Louis ME. Principles and Practice of Public Health Surveillance. Oxford, England: Oxford University Press; Aug 26, 2010.
28. Oehmke J, Oehmke T, Singh L, Post L. Dynamic panel estimate-based health surveillance of SARS-CoV-2 infection rates to inform public health policy: model development and validation. J Med Internet Res 2020 Sep 22;22(9):e20924 [FREE Full text] [doi: 10.2196/20924] [Medline: 32915762]
29. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
30. Oehmke J, Moss CB, Singh L, Oehmke TB, Post LA. Dynamic panel surveillance of COVID-19 transmission in the United States to inform health policy: observational statistical study. J Med Internet Res 2020 Oct 05;22(10):e21955 [FREE Full text] [doi: 10.2196/21955] [Medline: 32924962]
31. Wells CR, Townsend JP, Pandey A, Krieger G, Singer BH, McDonald RH, et al. Optimal COVID-19 quarantine and testing strategies. medRxiv Preprint posted online November 8, 2020. [doi: 10.1101/2020.10.27.20211631] [Medline: 33173923]
32. Gershon D, Lipton A, Levine H. Managing covid-19 pandemic without destructing the economy. arXiv Preprint posted online April 21, 2020. [FREE Full text]
33. Feyisa HL. The World Economy at COVID-19 quarantine: contemporary review. Int J Economics Finance Manage Sci 2020;8(2):63. [doi: 10.11648/j.ijefm.20200802.11]
34. Khan N, Fahad S, Faisal S, Naushad M. Quarantine role in the control of corona virus in the world and its impact on the world economy. SSRN J 2020:1. [doi: 10.2139/ssrn.3556940]
35. Kupferschmidt K. Why do some COVID-19 patients infect many others, whereas most don't spread the virus at all? Science 2020 May 20:1. [doi: 10.1126/science.abc8931]
36. Tufekci Z. This overlooked variable is the key to the pandemic It's not R. The Atlantic. 2020 Sep 30. URL: https://www.theatlantic.com/health/archive/2020/09/k-overlooked-variable-driving-pandemic/616548/ [accessed 2020-09-30]
37. Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res 2020;5:67 [FREE Full text] [doi: 10.12688/wellcomeopenres.15842.3] [Medline: 32685698]
38. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, et al. Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. arXiv Preprint posted online May 27, 2020. [FREE Full text]
39. The COVID Tracking Project. 2020. URL: https://covidtracking.com/ [accessed 2020-11-15]
40. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
41. COVID-19 hospitalization and death by race/ethnicity. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html [accessed 2020-10-03]
42. Sharma G, Volgman AS, Michos ED. Sex differences in mortality from COVID-19 pandemic: are men vulnerable and women protected? JACC Case Rep 2020 Jul 15;2(9):1407-1410 [FREE Full text] [doi: 10.1016/j.jaccas.2020.04.027] [Medline: 32373791]
43. Klein SL, Dhakal S, Ursin RL, Deshpande S, Sandberg K, Mauvais-Jarvis F. Biological sex impacts COVID-19 outcomes. PLoS Pathog 2020 Jun;16(6):e1008570 [FREE Full text] [doi: 10.1371/journal.ppat.1008570] [Medline: 32569293]
44. The COVID Tracking Project. GitHub. 2020. URL: https://github.com/COVID19Tracking [accessed 2020-11-15]
45. Sigmund M, Ferstl R. Panel vector autoregression in R with the package panelvar. Q Rev Economics Finance 2019 Jan:1. [doi: 10.1016/j.qref.2019.01.001]
46. Croissant Y. Linear models for panel data. The Comprehensive R Archive Network. 2020 Oct 08. URL: https://cran.r-project.org/web/packages/plm/plm.pdf [accessed 2020-10-15]
47. Tasnim S, Hossain MM, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health 2020 May;53(3):171-174. [doi: 10.3961/jpmph.20.094] [Medline: 32498140]
48. Pennycook G, McPhetres J, Zhang Y, Lu J, Rand D. Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci 2020 Jul;31(7):770-780 [FREE Full text] [doi: 10.1177/0956797620939054] [Medline: 32603243]

49. Sharma K, Seo S, Meng C, Rambhatla S, Liu Y. Covid-19 on social media: analyzing misinformation in Twitter conversations. arXiv Preprint posted online March 26, 2020. [FREE Full text]
50. Enberg J. How COVID-19 is testing social media's ability to fight misinformation. Insider Intelligence eMarketer. 2020 Mar 18. URL: https://www.emarketer.com/content/how-covid-19-is-testing-social-medias-ability-to-fight-misinformation [accessed 2020-10-01]
51. Gupta L, Gasparyan AY, Misra DP, Agarwal V, Zimba O, Yessirkepov M. Information and misinformation on COVID-19: a cross-sectional survey study. J Korean Med Sci 2020 Jul 13;35(27):e256 [FREE Full text] [doi: 10.3346/jkms.2020.35.e256] [Medline: 32657090]
52. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
53. Singh L, Bansal S, Bode L, Budak C, Chi G, Kawintiranon K, et al. A first look at COVID-19 information and misinformation sharing on Twitter. arXiv Preprint posted online March 31, 2020. [FREE Full text]
54. Brennen JS, Simon F, Howard PN, Nielsen RK. Types, sources, and claims of Covid-19 misinformation. Reuters Institute for the Study of Journalism. 2020 Apr 07. URL: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [accessed 2020-10-01]
55. Smith GD, Ng F, Ho Cheung Li W. COVID-19: emerging compassion, courage and resilience in the face of misinformation and adversity. J Clin Nurs 2020 May;29(9-10):1425-1428 [FREE Full text] [doi: 10.1111/jocn.15231] [Medline: 32155302]
56. Garrett L. COVID-19: the medium is the message. Lancet 2020 Mar 21;395(10228):942-943 [FREE Full text] [doi: 10.1016/S0140-6736(20)30600-0] [Medline: 32171075]
57. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 2020 Aug;584(7820):262-267. [doi: 10.1038/s41586-020-2404-8] [Medline: 32512578]
58. Haffajee RL, Mello MM. Thinking globally, acting locally - the U.S. response to Covid-19. N Engl J Med 2020 May 28;382(22):e75. [doi: 10.1056/NEJMp2006740] [Medline: 32240580]
59. Gostin LO, Hodge JG, Wiley LF. Presidential powers and response to COVID-19. JAMA 2020 Apr 28;323(16):1547-1548. [doi: 10.1001/jama.2020.4335] [Medline: 32186661]
60. Capano G, Howlett M, Jarvis DS, Ramesh M, Goyal N. Mobilizing policy (in)capacity to fight COVID-19: understanding variations in state responses. Policy Soc 2020 Jul 03;39(3):285-308. [doi: 10.1080/14494035.2020.1787628]
61. Walther C. Development and Connection in the Time of COVID-19 Corona's Call for Conscious Choices. New York, NY: Palgrave Macmillan, Cham; 2020.
62. Dunford M, Qi B. Global reset: COVID-19, systemic rivalry and the global order. Res Globalization 2020 Dec;2:100021. [doi: 10.1016/j.resglo.2020.100021]
63. Mallinson DJ. Cooperation and conflict in state and local innovation during COVID-19. Am Rev Public Adm 2020 Jul 13;50(6-7):543-550. [doi: 10.1177/0275074020941699]
64. Cronin CJ, Evans WN. Private precaution and public restrictions: what drives social distancing and industry foot traffic in the COVID-19 era? National Bureau of Economic Research. 2020. URL: https://www.nber.org/system/files/working_papers/w27531/w27531.pdf
65. Maxeiner JR. America's covid-19 preexisting vulnerability: a government of men, not laws. Theory Pract Legislation 2020 Sep 16;8(1-2):213-235. [doi: 10.1080/20508840.2020.1809173]
66. Lyu W, Wehby GL. Community use Of face masks and COVID-19: evidence from a natural experiment of state mandates in the US. Health Aff (Millwood) 2020 Aug;39(8):1419-1425. [doi: 10.1377/hlthaff.2020.00818] [Medline: 32543923]
67. Peirlinck M, Linka K, Sahli Costabal F, Kuhl E. Outbreak dynamics of COVID-19 in China and the United States. Biomech Model Mechanobiol 2020 Dec;19(6):2179-2193 [FREE Full text] [doi: 10.1007/s10237-020-01332-5] [Medline: 32342242]
68. Mckinley J, Ferré-Sadurní L. Cuomo rejects N.Y.C.'s shutdown plan for virus spike, but may offer own. The New York Times. 2020 Oct 05. URL: https://www.nytimes.com/2020/10/05/nyregion/cuomo-shutdown-coronavirus.html [accessed 2020-10-06]
69. Johnson S. Wisconsin Supreme Court overturns the state's stay-at-home orders. NPR. 2020 May 14. URL: https://www.npr.org/2020/05/14/855855749/wisconsin-supreme-court-overturns-the-states-stay-at-home-orders [accessed 2020-10-08]
70. Eggert D, White E. Michigan Governor's virus powers upended with court ruling. US News & World Report. 2020 Oct 02. URL: https://www.usnews.com/news/best-states/michigan/articles/2020-10-02/539k-signatures-submitted-to-repeal-whitmer-emergency-powers [accessed 2020-10-08]
71. Pettengill M, McAdam A. Can we test our way out of the COVID-19 pandemic? J Clin Microbiol 2020 Oct 21;58(11):e02225-e02220. [doi: 10.1128/JCM.02225-20] [Medline: 32843531]
72. Wilson SR, Hans RF, Denn MP, Dent C, Baker III C, McKee LE. COVID-19: New York and other Northeast Council states take phased approach to reopening economy. DLA Piper. 2020 May 06. URL: https://www.dlapiper.com/en/us/insights/publications/2020/05/new-york-and-other-northeast-council-states-take-phased-approach-to-reopening-economy/ [accessed 2020-10-08]

73. Dr. Bell: “Data shows that wearing face masks works”; latest COVID-19 update. SCDHEC. 2020 Aug 12. URL: https://scdhec.gov/news-releases/dr-bell-data-shows-wearing-face-masks-works-latest-covid-19-update-august-12-2020 [accessed 2020-08-15]
74. COVID-19 policy response (CPR) portal. International Food Research Policy Institute. 2020. URL: https://www.ifpri.org/project/covid-19-policy-response-cpr-portal [accessed 2020-10-15]
75. Schmitz R. Behind Germany's relatively low COVID-19 fatality rate. NPR. 2020 Mar 25. URL: https://www.npr.org/2020/03/25/821591044/behind-germanys-relatively-low-covid-19-fatality-rate [accessed 2020-10-15]
76. Leeb RT, Price S, Sliwa S, Kimball A, Szucs L, Caruso E, et al. COVID-19 trends among school-aged children - United States, March 1-September 19, 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 02;69(39):1410-1415. [doi: 10.15585/mmwr.mm6939e2] [Medline: 33001869]
77. Zimmermann P, Curtis N. Coronavirus infections in children including COVID-19: an overview of the epidemiology, clinical features, diagnosis, treatment and prevention options in children. Pediatr Infect Dis J 2020 May;39(5):355-368 [FREE Full text] [doi: 10.1097/INF.0000000000002660] [Medline: 32310621]
78. Cases, reporting and tracing. Show Me Renewal. 2020. URL: https://renewal.missouri.edu/student-cases/ [accessed 2020-10-15]
79. Rittenberg A, Dinrich H. Big Ten football to resume weekend of Oct. 24. ESPN. 2020 Sep 16. URL: https://www.espn.com/college-football/story/_/id/29897305/sources-big-ten-announce-october-return [accessed 2020-10-03]
80. Kelly A. Wisconsin records highest single-day number of new COVID-19 cases. The Hill. 2020 Oct 01. URL: https://thehill.com/changing-america/well-being/prevention-cures/519204-wisconsin-records-highest-single-day-number-of [accessed 2020-10-03]
81. Alberty E. Utah coronavirus cases up 1,107 Friday, as death toll moves closer to 500. The Salt Lake Tribune. 2020. URL: https://www.sltrib.com/news/2020/10/02/utah-coronavirus-cases-up/ [accessed 2020-10-03]
82. Guzman J. Fauci says new coronavirus actions by Florida governor are 'very concerning'. The Hill. 2020 Sep 28. URL: https://thehill.com/changing-america/well-being/prevention-cures/518577-fauci-says-new-coronavirus-actions-by-florida [accessed 2020-10-02]

Edited by G Eysenbach; submitted 15.10.20; peer-reviewed by B Rosales, P Banik, A Carnahan; comments to author 31.10.20; revised version received 19.11.20; accepted 20.11.20; published 03.12.20.

Please cite as:
Post LA, Issa TZ, Boctor MJ, Moss CB, Murphy RL, Ison MG, Achenbach CJ, Resnick D, Singh LN, White J, Faber JMM, Culler K, Brandt CA, Oehmke JF
Dynamic Public Health Surveillance to Track and Mitigate the US COVID-19 Epidemic: Longitudinal Trend Analysis Study
J Med Internet Res 2020;22(12):e24286
URL: https://www.jmir.org/2020/12/e24286
doi:10.2196/24286
PMID:33216726

Original Paper

Detection of Hate Speech in COVID-19–Related Tweets in the Arab Region: Deep Learning and Topic Modeling Approach

Raghad Alshalan[1*], MSc; Hend Al-Khalifa[1*], Prof Dr; Duaa Alsaeed[1*], PhD; Heyam Al-Baity[1*], PhD; Shahad Alshalan[1*], MSc

King Saud University, Riyadh, Saudi Arabia
[*]all authors contributed equally

Corresponding Author:
Hend Al-Khalifa, Prof Dr
King Saud University
Riyadh 11451
Riyadh,
Saudi Arabia
Phone: 966 504426816
Email: hend.alkhalifa@gmail.com

Abstract

Background: The massive scale of social media platforms requires an automatic solution for detecting hate speech. These automatic solutions will help reduce the need for manual analysis of content. Most previous literature has cast the hate speech detection problem as a supervised text classification task using classical machine learning methods or, more recently, deep learning methods. However, work investigating this problem in Arabic cyberspace is still limited compared to the published work on English text.

Objective: This study aims to identify hate speech related to the COVID-19 pandemic posted by Twitter users in the Arab region and to discover the main issues discussed in tweets containing hate speech.

Methods: We used the ArCOV-19 dataset, an ongoing collection of Arabic tweets related to COVID-19, starting from January 27, 2020. Tweets were analyzed for hate speech using a pretrained convolutional neural network (CNN) model; each tweet was given a score between 0 and 1, with 1 being the most hateful text. We also used nonnegative matrix factorization to discover the main issues and topics discussed in hate tweets.

Results: The analysis of hate speech in Twitter data in the Arab region identified that the number of non–hate tweets greatly exceeded the number of hate tweets, where the percentage of hate tweets among COVID-19 related tweets was 3.2% (11,743/547,554). The analysis also revealed that the majority of hate tweets (8385/11,743, 71.4%) contained a low level of hate based on the score provided by the CNN. This study identified Saudi Arabia as the Arab country from which the most COVID-19 hate tweets originated during the pandemic. Furthermore, we showed that the largest number of hate tweets appeared during the time period of March 1-30, 2020, representing 51.9% of all hate tweets (6095/11,743). Contrary to what was anticipated, in the Arab region, it was found that the spread of COVID-19–related hate speech on Twitter was weakly related with the dissemination of the pandemic based on the Pearson correlation coefficient ($r=0.1982$, $P=.50$). The study also identified the commonly discussed topics in hate tweets during the pandemic. Analysis of the 7 extracted topics showed that 6 of the 7 identified topics were related to hate speech against China and Iran. Arab users also discussed topics related to political conflicts in the Arab region during the COVID-19 pandemic.

Conclusions: The COVID-19 pandemic poses serious public health challenges to nations worldwide. During the COVID-19 pandemic, frequent use of social media can contribute to the spread of hate speech. Hate speech on the web can have a negative impact on society, and hate speech may have a direct correlation with real hate crimes, which increases the threat associated with being targeted by hate speech and abusive language. This study is the first to analyze hate speech in the context of Arabic COVID-19–related tweets in the Arab region.

(J Med Internet Res 2020;22(12):e22609) doi:10.2196/22609

KEYWORDS

COVID-19; coronavirus; Twitter; hate speech; social network analysis; social media; public health; pandemic; deep learning; non-negative matrix factorization; NMF; convolutional neural network; CNN

Introduction

Social media platforms such as Twitter provide valuable venues for information sharing, communication, and knowledge production. However, these platforms have also been increasingly exploited for the propagation of hate speech. Hate speech can be generally defined as language that aims to target a specific group on the basis of characteristics such as ethnic origin, religion, or gender [1].

Hate speech on the internet is a complex concept with a wide spectrum of targets, forms, and other related concepts [2,3]. While there is no formal definition of hate speech, there is general agreement among scholars and service providers to define it as any language that attacks a person or a group based on a characteristic such as race, color, ethnicity, gender, sexual orientation, nationality, or religion [3]. This type of discriminatory and hateful speech can have a destructive impact on society, as it threatens the culture of coexistence and unity. It is also evident that hate speech has a strong connection with actual hate crimes [3,4], which increases the risks associated with being targeted by hate speech and abusive language.

In mid-February 2020, the COVID-19 virus pandemic started to penetrate the Arab region physically and virtually, and Arab social media users started talking about the spread of the disease using different platforms. The pandemic became a trending topic that was visible to all users on Twitter within the Arab region. Since then, multiple voices have been raised across social media using the COVID-19 pandemic as a vehicle for spreading hate speech.

The use of social media during pandemics for a variety of public health purposes has been demonstrated in a growing body of literature. One systematic review on social media and emerging infectious diseases identified three major approaches, namely assessment of public opinion, social media use by organizations, and evaluation of information accuracy [4]. Another systematic review on studies that primarily use web-based social networks for pandemic detection and tracking found that social networks have rich information that can be utilized to develop an almost real-time pandemic surveillance system [5]. Other studies have used Twitter and other sources of data to build a surveillance system [6] and develop automatic methods for quantifying the scientific quality and sensationalism of individual news records during the pandemic [7].

During the unfolding COVID-19 pandemic, many research studies have focused on Twitter to characterize the impact of the pandemic on public responses and behaviors. One of these studies analyzed the main topics discussed by Twitter users during the crisis by leveraging techniques such as latent Dirichlet allocation (LDA) and sentiment analysis [8]. Another study found that frequent use of social media during the pandemic contributed to information overload, which had a significant impact on individuals' coping perceptions [9]. In addition, web-based social question-and-answer forums have been analyzed to identify topic communities and asses the appropriateness of the answers during the early stage of the COVID-19 outbreak [10]. Other studies have analyzed Twitter data to understand the impact of COVID-19 on specific public attitudes and behaviors, such as xenophobia [11] and the spread of the 5G conspiracy theory [12]. As discussed in [11], the pandemic has triggered discrimination and stigma toward specific groups on social media. This phenomenon was further explored in [13], with a specific focus on building an automatic tool for detecting hate speech in Spanish tweets related to newspaper articles about the COVID-19 pandemic. The preliminary results showed that 9% of the tweets contained hate speech.

As can be seen, the massive scale of social media platforms requires an automatic solution for detecting hate speech. Such solutions will help reduce the need for manual analysis of content. Most previous literature has cast the hate speech detection problem as a supervised text classification task using classical machine learning methods or, more recently, deep learning methods. However, studies investigating this problem in Arabic cyberspace are still limited compared to the published work on English text.

The aim of this study is to automatically identify hate speech posted by Arabic-speaking Twitter users related to the COVID-19 pandemic using a deep learning approach. We opted to analyze the spread of hate speech on the Twitter platform given its increasing popularity among Arab users [14]. According to [15], nearly 9 out of 10 Arab adults use the internet, and 22% of them use the Twitter platform actively. In Saudi Arabia alone, there are more than 10 million active Twitter users, akin to almost 38% of the population.

Our research objectives are to answer the following research questions: (1) In which Arab country were the most COVID-19–related hate tweets posted during the pandemic? (2) When were the highest number of hate tweets posted in the Arab region during the pandemic? (3) Does the increase in the number of COVID-19 cases or deaths coincide with an increase in the number of hate tweets? and (4) What are the main topics and issues being discussed in hate tweets during the COVID-19 pandemic?

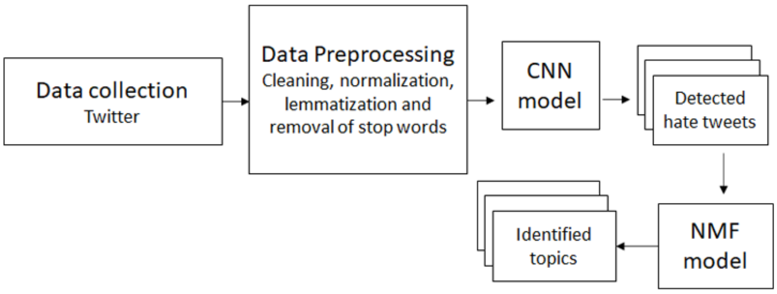
The rest of the paper is structured as follows: the Methods section presents our methods, including a description of the neural network models and the preprocessing steps. In the next sections, we provide the results and discuss them. Finally, we conclude our work and discuss future directions.

Methods

In this study, we aimed to automatically detect hate tweets posted by Twitter users in the Arab region during the COVID-19 pandemic using a deep learning approach and topic modeling.

The workflow of our methodology is depicted in Figure 1. In the following subsections, we describe each step in more detail.

Figure 1. Methodology workflow. CNN: convolutional neural network; NMF: utilized nonnegative matrix factorization.



Data Collection

We used the ArCOV-19 [16] dataset, which is an ongoing collection of Arabic tweets related to COVID-19 starting from January 27, 2020. The tweets were collected using the Twitter search application programming interface (API) to obtain the most popular tweets using different queries, such as simple keywords (eg, *Corona*), hashtags (eg, *#Corona*), or phrases (eg, *novel Coronavirus*). The search was customized to exclude retweets and non-Arabic tweets. At the time of writing this paper, the ArCOV-19 dataset contained almost 1 million popular tweets (ie, tweets that received the maximum social interactions and impressions) collected from January 27 to April 30, 2020. In compliance with the Twitter content redistribution policy, the authors of the ArCOV-19 dataset only released the tweet IDs. Therefore, we recollected the full tweet objects using the *Twarc* library [17], which is a Python library for retrieving Tweets as JavaScript Object Notation (JSON) objects using the Twitter API based on the given IDs. We started the preprocessing pipeline by removing retweets and duplicate tweets that had identical content to reduce the effect of malicious tweets posted by social bots. After that, and due to the inaccessibility of some tweets (deleted tweets or deactivated accounts), the total number of tweets we retrieved was reduced to 975,316.

Data Preprocessing

The retrieved JSON object for each tweet is a mix of many attributes. For this study, we retained the following attributes: tweet ID, tweet text, creation time, user ID, and location. The location field represents the location profile metadata as declared by the Twitter user. However, this field contains free-form text entered manually by the user; hence, this field can be noisy, as users can type anything. Therefore, we needed to resolve the text to an exact location (in our case, a country).

The procedure to extract the country from the location text was as follows. First, for each country in the Arab region, we prepared a list that contains the country names and all possible city names derived from GeoNames [18], a geographical database that contains over 25 million geographical names. After that, each list was translated into Arabic and automatically extended to include different morphological variant forms of the same word. Also, each list was manually modified to include more informal terms and forms that Twitter users frequently use to refer to their countries or cities. For example, Emirati users might type “Dar Zayed” in their location field instead of “United Arab Emirates” or “UAE”, and Kuwaiti users usually write “Q8” instead of “Kuwait”. Also, to minimize the mismatching problem between the prepared list and the location field, we normalized the names in our lists and the location field to reduce orthographical variations; for example, different forms of ”” (“”, “”, and “”) were replaced with “”, while the letter “” was replaced with “” and the letter “” was replaced with “”. After that, we performed several cleaning and denoising steps on the lists and location fields, including the removal of punctuation marks, additional white spaces, and diacritics. Then, to map each text location to a specific country, we checked if at least one of the words in the location text appeared in the list for any country. If it did, the location was resolved to that specific country. For example, if a user typed “Riyadh”, “Najd”, “Jeddah”, or “Saudi”, we resolved this location to “Saudi Arabia.” Finally, we dropped all tweets that belonged to users with no detected location.

To prepare the tweets to be fed as inputs into the convolutional neural network (CNN) classification model and the nonnegative matrix factorization (NMF) model, we firstly applied several preprocessing steps to the text of the tweets. First, we removed punctuation, consecutive words, additional white spaces, and non-Arabic letters. We also replaced emojis with their descriptions for CNN inputs and removed them for the NMF model. We also normalized Twitter-specific tokens by converting the links and mentions to “URL” and “mention,” respectively. After that, various forms of the same word were lemmatized by converting them to the main word using the Farasa tool [19]; finally, we removed stop words from the text.

Hate Speech Detection Using the CNN Model

The processed tweets were analyzed for hate speech using the pretrained CNN model. The CNN model used in this study is based on our previous work [20] in which we addressed the problem of hate speech spread in the Arabic Twittersphere. In this study, a CNN was trained on almost 10,000 tweets that were manually labeled as hate tweets or non–hate tweets. The model contained five layers: an input layer (embedding layer), a convolution layer, which basically consists of 3 parallel convolution layers with different kernel sizes (2, 3, and 4), a pooling layer, a hidden dense layer, and a final output layer that assigned a “hate score” (ranging from 0 to 1) to each text. The results showed that the CNN model achieved the best performance (compared to other models), with an F1 score of 79% and accuracy of 83%. Therefore, we used it in this study.

The analysis was performed on the tweet texts, and each tweet was given a score between 0 and 1, with 1 being the most hateful text. We used the default threshold of 0.5 (or 50%) to classify a prediction with a probability of 0.5 or more as a hate tweet and any value less than 0.5 as a non–hate tweet. We further classified hate prediction into three levels based on score: low (0.50-0.67), average (0.68-0.85), and high (0.86-1.00). The hate tweets were then categorized into three levels (low, colored yellow; average, colored orange; and high, colored red) based on the detected hate scores.

Topic Modeling Using NMF

NMF is an unsupervised approach for reducing the dimensionality of nonnegative matrices [21], and it has been widely used to discover the underlying relationships between texts and identify latent topics [22,23]. In this study, we opted to employ NMF to discover the semantic structure and discussed topics within the set of hate tweets. We particularly chose NMF given its ability to produce more coherent topics than other popular topic modeling techniques such as the LDA model, as shown in this study. We used the sklearn library [24] to implement the NMF model. To apply NMF, we first preprocessed the tweets as described earlier. Then, we created the features by transforming the processed tweets into term frequency–inverse document frequency vectors (unigrams and bigrams).

Results

Model Results

As mentioned, the first step in the hate speech analysis was to analyze the processed tweets using the pretrained CNN model. Table 1 presents the total numbers of detected hate tweets and non–hate tweets in the Arab region, along with the three levels of hate (low, average, and high). It can be observed in Table 1 that 535,811 of the 547,554 collected tweets related to the COVID-19 pandemic (97.8%) could be classified as non–hate tweets, while 11,743 (3.2%) were classified as hate tweets. Generally, most of the hate tweets fell in the low level range (8385/11,743, 71.4%). On the other hand, only 2.89% of hate tweets (340/11,743) were categorized as high level.

Table 1. Total numbers of hate tweets and non–hate tweets in the Arab region during the period of study (N=547,554).

Type of tweet	Number of tweets, n (%)
Non–hate tweets	535,811 (97.8)
Hate tweets	11,743 (3.2)
Hate levels[a] (n=11,743)	
Low	8385 (71.4)
Average	3018 (25.7)
High	340 (2.9)

[a] Based on scores assigned by the convolutional neural network model (low: 0.50-0.67; average: 0.68-0.85; high: 0.86-1.00).

Descriptive Statistical Analysis

To obtain more insight into the obtained results, we analyzed the hate tweets from two different aspects: by country and by time period. In addition, we obtained the WHO COVID-19 statistics (number of cases and deaths) for each country and during the aforementioned time periods.

By Country

The map in Figure 2 displays the number of hate tweets and their distribution across the Arab countries, where the number of hate tweets is shown in red, the number of total tweets is shown in black, and a dark color depicts a high number of hate tweets in that country.

Figure 2. Number of hate tweets (red) and total tweets (black) in each Arab country, where a darker color depicts a higher number of hate tweets in that country (MA: Morocco, MR: Mauritania, DZ: Algeria, TN: Tunisia, LY: Libya, EG: Egypt, JO: Jordan, LB: Lebanon, SY: Syria, IQ: Iraq, SA: Saudi Arabia ,YE: Yemen, KW: Kuwait, QA: Qatar, AE: United Aran Emirates, OM: Oman).

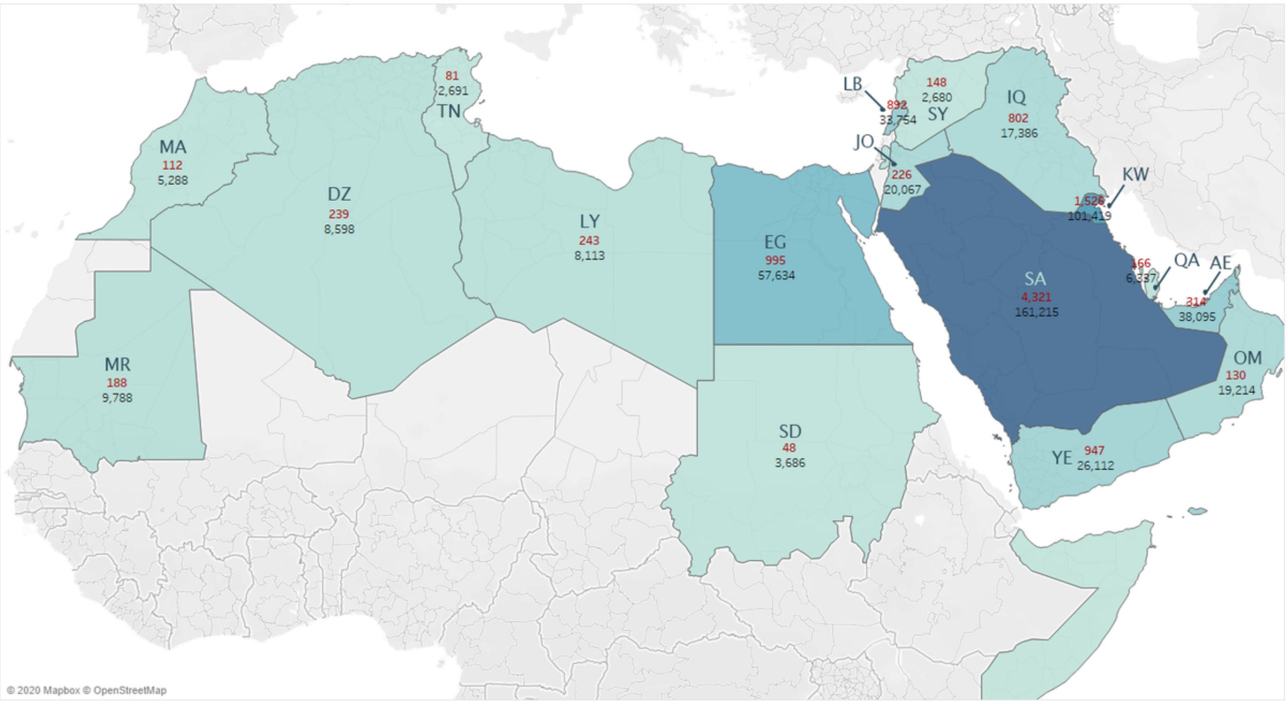


Table 2 compares the number of hate tweets in Arab countries and the distribution of hate tweets with different hate levels (low, average, and high). Out of the 22 Arab countries, the 10 most Twitter-active countries with respect to the number of COVID-19–related tweets are shown in the table. The results of the other 12 countries are combined in one column, titled “Other Arab countries.” In addition, Table 2 presents COVID-19 statistics (number of cases and deaths) for each country. It can be observed from Table 2 that country-wise, most COVID-19 tweets are of the non-hate class, and the percentages of hate tweets are considerably low. For instance, Saudi Arabia, which is the most Twitter-active Arab country as stated in the Digital 2020 Global Report [25], has the highest number of COVID-19–related tweets (165,536 tweets), of which only 4321 (2.61%) were classified as hate tweets.

In line with our previous observation, where most of the hate tweets were at the low level of hate, we found that the largest numbers of hate tweets were at the low hate level. For example, 2813 of the 4321 hate tweets in Saudi Arabia (65.10%) fell in the low level range, while 200 (4.63%) were at a high level. Similar results can be seen when considering Mauritania, which has the smallest number of COVID-19–related tweets among the top 10 Arab countries. 81.4% of hate tweets (153/188) in Mauritania are at the low hate level, and only 1.1% (2/188) are at a high level.

Table 2. Statistics of COVID-19–related hate tweets posted in Arab countries.

Variable	Country										
	Saudi Arabia	Kuwait	Egypt	UAE[a]	Lebanon	Yemen	Jordan	Oman	Iraq	Mauritania	Other Arab countries[b]
Population (million)	34	4	102	10	7	30	10	5	40	5	187
Posted tweets (N=535,811), n (%)	165,536 (24.2)	102,945 (15.0)	58,629 (8.6)	38,409 (5.6)	34,646 (5.1)	27,059 (4.0)	20,293 (3.0)	19,344 (2.8)	18,188 (2.7)	9976 (1.5)	52,529 (7.1)
Hate tweets (n=11,743), n (%)	4321 (2.6)	1526 (1.5)	995 (1.7)	314 (0.8)	892 (2.6)	947 (3.5)	226 (1.1)	130 (0.7)	802 (4.4)	188 (1.9)	1402 (2.7)
Hate levels of tweets[c]											
Low (n=8385), n (%)	2813 (65.1)	1153 (75.6)	747 (75.0)	245 (78.0)	700 (78.5)	640 (67.5)	160 (70.8)	104 (80.0)	624 (77.8)	153 (81.4)	1046 (74.6)
Average (n=3018), n (%)	1308 (30.3)	347 (22.7)	224 (22.5)	65 (20.7)	184 (20.6)	277 (29.2)	62 (27.4)	21 (16.6)	168 (20.9)	33 (17.6)	329 (23.5)
High (n=340), n (%)	200 (4.6)	26 (1.7)	24 (2.4)	4 (1.3)	8 (0.9)	30 (3.2)	4 (1.8)	5 (3.8)	10 (1.2)	2 (1.1)	27 (1.9)
Average hate level score	0.643	0.613	0.617	0.606	0.605	0.634	0.623	0.602	0.610	0.602	0.623
COVID-19 statistics (n)											
Cases (n=78,037)	22753	4024	5537	12481	725	6	453	2348	2003	7	27700
Deaths (n=1593)	162	26	392	105	24	0	8	11	92	1	772

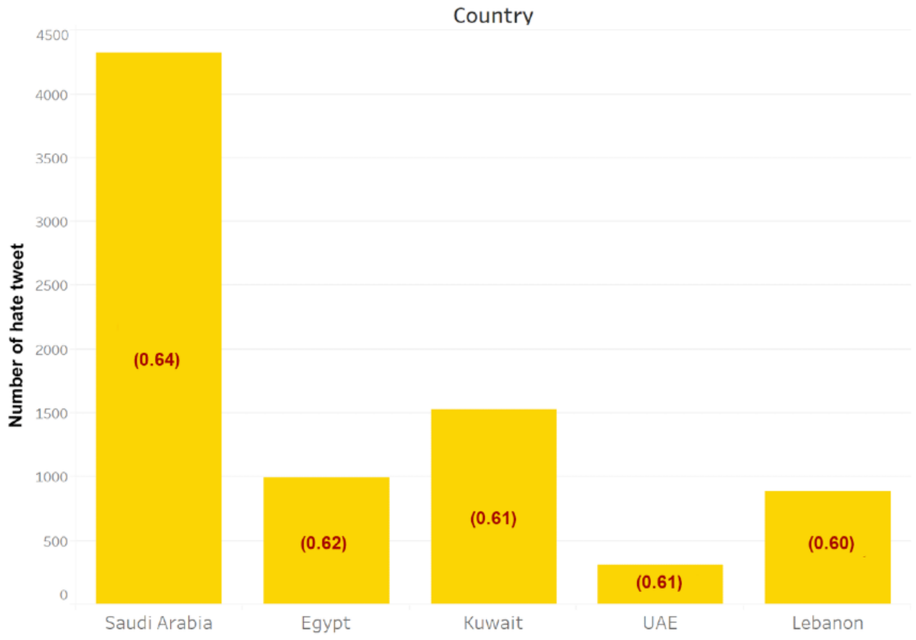
[a] UAE: United Arab Emirates.

[b] Other Arab countries: this column combines the results of the remaining Arab countries (Palestine, Algeria, Libya, Bahrain, Morocco, Qatar, Sudan, Syria, Tunisia, Comoros, and Somalia).

[c] Based on scores assigned by the convolutional neural network model (low: 0.50-0.67; average: 0.68-0.85; high: 0.86-1.00).

Figure 3 illustrates the statistics of hate tweets for the top five Arab countries with respect to number of COVID-19–related tweets, showing the average hate level of all hate tweets in red. It can be seen from Figure 3 that average hate level for the countries shown falls in the low hate range, which is coded in yellow. This strongly supports our first finding that the majority of hate tweets do not contain a high level of hate.

Figure 3. Number of COVID-19–related hate tweets per country with the average hate level scores in brackets (low: 0.50-0.67; average: 0.68-0.85; high: 0.86-1.00). UAE: United Arab Emirates.



By Time Period

The ArCOV-19 [15] data used in our study cover the period of January 27 to April 30, 2020. This period was divided into three specific time durations: January 27 to February 29, March 1-30, and April 1-30. Table 3 includes the number of COVID-19–related tweets in the Arab region during each time period, together with the number of hate tweets and their distribution in the three different hate levels. It also presents the number of COVID-19 cases and deaths in each time period.

Table 3. Statistics of COVID-19–related hate tweets per time period (N=547,554).

Variable	Time period (2020)		
	January 27-February 29	March 1-30	April 1-30
Total tweets (N=547,554), n (%)	118,991 (21.7)	253,806 (46.4)	174,757 (31.9)
Hate tweets (n=11,743), n (%)	3014 (25.7)	6095 (51.9)	2634 (22.4)
Hate levels of tweets[a]			
Low	2198 (72.9)	4300 (70.5)	1887 (71.6)
Average	741 (24.6)	1617 (26.5)	660 (25.1)
High	75 (2.5)	178 (2.9)	87 (3.3)
Average hate level score	0.622	0.628	0.625
COVID-19 statistics (n)			
Cases	133	7447	70,092
Deaths	0	202	1350

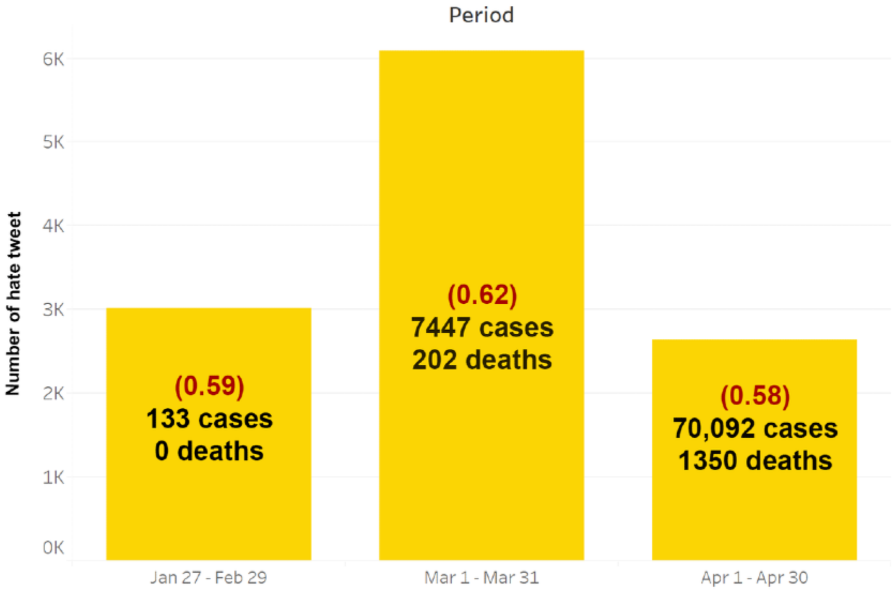
[a] Based on scores assigned by the convolutional neural network model (low: 0.50-0.67; average: 0.68-0.85; high: 0.86-1.00).

From Table 3, it can be clearly noted that after the end of the first period (January 27 to February 29), the number of COVID-19–related tweets spiked, with an increase of 113.3% compared to the first period. However, the number then decreases by 31.1% in the third period (April 1-30). Likewise, the number of hate tweets increased by 102.2% in the second period, then decreased by 56.8% in the third period. In contrast, the number of identified COVID-19 cases continued to grow during all three periods.

Looking at the levels of hate tweets per time period, it can be seen that although the number of hate tweets falling in the high level range is extremely small compared to those in the low level range, the high-level hate tweets have a tendency to increase with increasing number of COVID-19 cases and deaths, showing percentages of 2.49% (75/3014), 2.9% (178/6095), and 3.3% (87/2634) for the first, second, and third periods, respectively. On the other hand, the low and average hate levels do not follow the same behavior.

Figure 4 illustrates the statistics of the number of hate tweets per time period and the numbers of COVID-19 cases and deaths together with the average hate levels of all hate tweets (shown in red). It can be observed from Figure 4 that the average hate levels for the three time periods are in the low-level range, which is in line with our initial finding that the majority of hate tweets do not reach the high level of hate.

Figure 4. Numbers of hate tweets and numbers of COVID-19 cases and deaths per time period with the average hate level scores in brackets (low: 0.50-0.67; average: 0.68-0.85; high: 0.86-1.00).



From the previous results, it can be seen that there is a weak link between the number of hate tweets and the spread of COVID-19 per country or time period. This is because the number of hate tweets does not follow the same pattern of the number of COVID-19 cases, as illustrated in Table 2 and Table 3. To investigate this further, we tested the correlations between the number of COVID-19–related hate tweets and the number of COVID-19 cases in the data of 21 Arab countries during the 3 time periods under study. Using the Pearson correlation test [26], we obtained an *r* value of 0.1982 at *P*=.50. Although *r* is technically positive, the relationship between the number of hate tweets and the number of COVID-19 cases is weak (the nearer the value of *r* to zero, the weaker the relationship [26]).

NFM Results

To understand the main topics discussed by Arab users during the pandemic, we employed NFM as discussed earlier in this paper. We extracted 7 topics from the hate tweets. The distributions and examples of the top terms associated with each topic are shown in Table 4. As shown in Table 4, the most relevant terms for Topic 1 were terms that were usually used to criticize China as a country for being the source of SARS-CoV-2, the virus that cause COVID-19. Within the same theme, tweets in Topic 4 expressed hate speech toward Chinese people and their eating habits as being responsible for spreading COVID-19 worldwide. China was also the target of attack in Topic 5, in which many users expressed their belief that COVID-19 served as a “divine punishment against China” because of the Uyghur-Chinese conflict.

Table 4. Identified topics in hate tweets and examples of the most common words in each topic.

Topic number	Main theme	Examples of the top unigrams and bigrams	Distribution
1	China and COVID-19	Cursed, China, curse China, Curses on, life, new	4.3%
2	Iran as a source of COVID-19	Export, Gulf, country, Terrorism, disease, spread	5.07%
3	Saudi citizens visiting Iran	Saudi Arabia, Saudi, Bahrain, citizen, travel, passport	9.44%
4	Chinese eating habits and COVID-19	Dog, animal, bat, eating, meat, snake	5.72%
5	China and Uighur	China, Muslim, Uyghur, Chinese, pig, punishment	7.38%
6	Iran regime	Regime, Mullahs, Iran, Iranian People, Tehran, outbreak	5.60%
7	General political tweets, conspiracies, COVID-19 as an exaggerated threat	Country, people, disease, party, Iraq, Egypt	62.45%

On the other hand, Topic #2 contained several tweets accusing Iran of deliberately spreading COVID-19 to the Gulf Cooperation Council (GCC) countries. Topic #6 is also about Iran and COVID-19; however, the tweets in this topic were more political, attacking the Iranian regime and its politics. Topic #3 included cases of hate speech against Saudi citizens who visited Iran during the pandemic as being responsible for the spread of COVID-19 in Saudi Arabia.

Finally, in Topic #7, we observed that most of the inspected tweets were written in a political context, where the most frequently mentioned countries were Iraq and Egypt. We also observed many tweets spreading conspiracy theories about the scale of the pandemic and its origin, and some tweets stated that COVID-19 is an exaggerated crisis compared to the current regional crises and conflicts.

Discussion

Principal Results

This study aimed to address four research questions: (1) In which Arab country were the most COVID-19 hate tweets posted during the COVID-19 pandemic? (2) When were the highest number of hate tweets posted in the Arab region during the pandemic? (3) Does the increase in the number of COVID-19 cases or deaths coincide with an increase in the number of hate tweets? and (4) What are the main topics and issues being discussed in hate tweets during the COVID-19 pandemic?

In regard to the Arab country from which the highest number of hate tweets related to the COVID-19 pandemic originated, Saudi Arabia, which has the highest number of COVID-19–related tweets (165,536 tweets), had the highest number of hate tweets (4321), as shown in Table 2. Therefore, Saudi Arabia is the country in which the most COVID-19 hate tweets were posted during the pandemic. This can be attributed to Saudi Arabia numbering in fourth place in the top 20 countries and the only Arab country in the top 20 based on the largest active Twitter audience worldwide as mentioned in the Digital 2020 Global Report [25].

However, the percentage of hate tweets in Saudi Arabia, which is 2.6% (4321/165,536), is not the highest in the Arab region. This indicates that the number of COVID-19–related tweets does not necessarily reflect the number of hate tweets (as highlighted in Table 2). This can be clearly seen in the case of Yemen, where 27,059 tweets were posted, 3.5% of which were classified as hate tweets (n=947). Yemen having the highest percentage of hate tweets can be linked to the political and security instability in Yemen due to the Yemen Civil War, which began in 2014, and its psychological effect on the citizens of this country.

In addressing the second research question, this paper identifies the time period during which the number of hate tweets was the highest in the Arab region. As shown in Table 3, the second period (March 1-30) has the highest number of hate tweets (6095 tweets), which represents 51.9% of the 11,743 hate tweets in the three-month period of January 27 to April 30, 2020. Despite the fact that the highest numbers of cases and deaths occurred in the third period (April 1-30), the number of hate tweets reached its maximum during the second period. There are two possible explanations for this finding. First, it may be related to the beginning of the spread of COVID-19 in the Arab region, which occurred in March 2020. This period witnessed the highest level of public attention, which can be seen through the high number of Twitter interactions during the pandemic. The other explanation is that people began to adjust to coping with the pandemic and health precautions in the third period, and the severity of their panic and anxiety decreased.

The third research question investigated whether the number of hate tweets coincided with the spread of COVID-19 in the Arab region. As clearly shown in Table 2 and Table 3, the spread of COVID-19 hate tweets does not follow the same trend of the increase or decrease in the number of COVID-19 cases or deaths in any Arab country or during the studied time period. This can be interpreted as a weak relationship between the two variables because the *r* value obtained from the Pearson correlation test was 0.1982 (*P*=.50) [26]. Moreover, it can be observed from Table 3 that most of the hate tweets appeared in the second period, while the numbers of COVID-19 cases and deaths are highest in the third period, with percentages of 90.2% and 87%, respectively.

Regarding the fourth research question, this study sought to identify the topics discussed by Arab Twitter users in hate tweets. We analyzed the 7 topics extracted from the NFM model, and as shown in Table 4, most of the identified topics revolved around two themes: China (topics 1, 4, and 5) and Iran (topics 2, 3, and 6). It can be clearly noted that the COVID-19 pandemic fueled nationalism and ethnic conflict among Arab Twitter users. This spike of hate speech against China and Iran (and Arab individuals who traveled to Iran) can be attributed to the fact that China and Iran were increasingly accused in the media of spreading SARS-CoV-2 worldwide and to Arab countries, respectively. The United Nations recently published a guidance note [27] on addressing and countering COVID-19–related hate speech. The note stated that the pandemic has given rise to a new wave of hate speech and discrimination, which our results clearly confirm. After manual inspection of a few sample tweets from the identified topics, we observed that most of the tweets incited hatred and blame against individuals or groups belonging to certain ethnicities or from certain countries. This type of COVID-19–related hate speech could entail serious consequences, making the targeted groups more vulnerable to violence and discrimination and exposing them to political and social exclusion, isolation, and stigmatization.

Limitations and Future Work

One strength of this study is that it automatically identified hate tweets posted during the COVID-19 pandemic in the Arab region using deep learning and topic modeling. The model was trained on data from similar domains with similar characteristics (ie, noisy tweets with a high variety of dialectal Arabic). The model was able to identify a total of 11,743 hate tweets, which allowed us to study the phenomenon of hate speech on Twitter. However, some limitations of our study do exist. One limitation is that the used model was trained on a dataset collected in a span of 6 months, and certain events and topics dominated social media at that time. This may have affected the model performance and prevented it from capturing new forms of hate, violence, and racism that could be exacerbated by the COVID-19 pandemic. Additional limitations are the short duration of the study and the chances of misclassification of countries due to self-reporting of users' locations.

In future research, a follow-up analysis could be conducted to compare the prevalence of hate with other psychological aspects, such as anxiety, fear, and depression, in Twitter data during the COVID-19 pandemic.

Conclusion

The COVID-19 pandemic poses a significant threat to public health in nations worldwide. This study conducted an analysis of hate speech in Twitter data in the Arab region using deep learning and topic modeling. The analysis revealed that the number of non–hate tweets greatly exceeds the number of hate tweets. We also found that the majority of hate tweets were at a low level of hate. This study has shown that in the Arab region, Yemen is the Arab country from which the highest percentage of COVID-19 hate tweets spread during the pandemic. Moreover, it was noted that the time period of March 1-30 showed the highest number of hate tweets, representing 51.9% of all hate tweets.

Contrary to what was expected, the results in this study showed that the number of hate tweets did not follow the same trend as the increase or decrease in the number of COVID-19 cases or deaths. In fact, it was found that the spread of COVID-19 hate speech on Twitter is weakly related to the dissemination of the pandemic based on the Pearson correlation test.

This study also explored the main topics discussed by Arab users in tweets that were identified as hatful. After analyzing the NMF model results, we found that the COVID-19 pandemic clearly triggered hate speech against China and Iran. We also found that most Arabic hate tweets posted during the pandemic propagated hate speech in a political context, which may have been triggered by the ongoing crisis and political tension in the Arab region.

Acknowledgments

This research was funded by Researchers Supporting Project (number RSP-2020/276), King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest

None declared.

References

1. Schmidt A, Wiegand M. A Survey on Hate Speech Detection using Natural Language Processing. In: Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media.: Association for Computational Linguistics; 2017 Presented at: Fifth International Workshop on Natural Language Processing for Social Media; April 2017; Valencia, Spain p. 1-10. [doi: 10.18653/v1/w17-1101]
2. Silva L, Mondal M, Correa D, Benevenuto F, Weber I. Analyzing the Targets of Hate in Online Social Media. ArXiv Preprint posted online on March 03, 2016. [FREE Full text]
3. Fortuna P, Nunes S. A Survey on Automatic Detection of Hate Speech in Text. ACM Comput Surv 2018 Sep 06;51(4):1-30. [doi: 10.1145/3232676]
4. Tang L, Bie B, Park S, Zhi D. Social media and outbreaks of emerging infectious diseases: A systematic review of literature. Am J Infect Control 2018 Sep;46(9):962-972. [doi: 10.1016/j.ajic.2018.02.010] [Medline: 29628293]
5. Al-garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: A systematic review. J Biomed Inform 2016 Aug;62:1-11. [doi: 10.1016/j.jbi.2016.05.005]
6. Jain VK, Kumar S. Effective surveillance and predictive mapping of mosquito-borne diseases using social media. J Comput Sci 2018 Mar;25:406-415. [doi: 10.1016/j.jocs.2017.07.003]
7. Hoffman SJ, Justicz V. Automatically quantifying the scientific quality and sensationalism of news records mentioning pandemics: validating a maximum entropy machine-learning model. J Clin Epidemiol 2016 Jul;75:47-55. [doi: 10.1016/j.jclinepi.2015.12.010]
8. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
9. Farooq A, Laato S, Islam AKMN. Impact of Online Information on Self-Isolation Intention During the COVID-19 Pandemic: Cross-Sectional Study. J Med Internet Res 2020 May 06;22(5):e19128 [FREE Full text] [doi: 10.2196/19128] [Medline: 32330115]
10. Jo W, Lee J, Park J, Kim Y. Online Information Exchange and Anxiety Spread in the Early Stage of the Novel Coronavirus (COVID-19) Outbreak in South Korea: Structural Topic Model and Network Analysis. J Med Internet Res 2020 Jun 02;22(6):e19455 [FREE Full text] [doi: 10.2196/19455] [Medline: 32463367]
11. Budhwani H, Sun R. Creating COVID-19 Stigma by Referencing the Novel Coronavirus as the "Chinese virus" on Twitter: Quantitative Analysis of Social Media Data. J Med Internet Res 2020 May 06;22(5):e19301 [FREE Full text] [doi: 10.2196/19301] [Medline: 32343669]
12. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
13. Cotik V, Debandi N, Luque F, Miguel P, Moro A, Perez J, et al. A study of Hate Speech in Social Media during the COVID-19 outbreak. 2020 Jul 01 Presented at: ACL 2020 NLP COVID-19 Workshop; July 9-19, 2020; Seattle, WA.
14. Twitter in the Arab Region. Arab Social Media Report. URL: https://www.arabsocialmediareport.com/Twitter/LineChart.aspx?&PriMenuID=18&CatID=25&mnu=Cat [accessed 2020-10-06]
15. Radcliffe D, Abuhmaid H. Social Media in the Middle East: 2019 in Review. SSRN Journal Preprint posted online on January 12, 2020. [doi: 10.2139/ssrn.3517916]
16. Haouari F, Hasanain M, Suwaileh R, Elsayed T. ArCOV-19: The First Arabic COVID-19 Twitter Dataset with Propagation Networks. ArXiv Preprint posted online on April 18, 2020. [FREE Full text]
17. twarc. GitHub. URL: https://github.com/DocNow/twarc [accessed 2020-12-03]
18. GeoNames. URL: https://www.geonames.org/ [accessed 2020-10-06]
19. Abdelali A, Darwish K, Durrani N, Mubarak H. Farasa: A Fast and Furious Segmenter for Arabic. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations.: Association for Computational Linguistics; 2016 Presented at: 2016 Conference of the North American Chapter of the Association for Computational Linguistics; June 12-17, 2016; San Diego, CA p. 11-16. [doi: 10.18653/v1/n16-3003]
20. Alshalan R, Al-Khalifa H. A Deep Learning Approach for Automatic Hate Speech Detection in the Saudi Twittersphere. Applied Sciences 2020 Dec 01;10(23):8614. [doi: 10.3390/app10238614]
21. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999 Oct 21;401(6755):788-791. [doi: 10.1038/44565] [Medline: 10548103]
22. Arora S, Ge R, Moitra A. Learning Topic Models - Going beyond SVD. In: arXivcs Internet. 2012 Apr 9 Presented at: 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science; October 20-23, 2012; New Brunswick, NJ. [doi: 10.1109/focs.2012.49]
23. Shi T, Kang K, Choo J, Reddy C. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In: WWW '18: Proceedings of the 2018 World Wide Web Conference. 2018 Apr Presented at: 2018 World Wide Web Conference; April 23-27, 2018; Lyons, France p. 1105-1114. [doi: 10.1145/3178876.3186009]
24. sklearn.decomposition.NMF. scikit learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF.html [accessed 2020-12-03]
25. The Global State of Digital 2020. Hootsuite. URL: https://hootsuite.com/pages/digital-2020 [accessed 2020-07-17]

26. Benesty J, Chen J, Huang Y, Cohen I. Pearson Correlation Coefficient. In: Noise Reduction in Speech Processing. Springer Topics in Signal Processing, vol 2. Berlin, Germany: Springer; 2009:1-4.
27. United Nations guidance note on addressing and countering COVID-19 related hate speech. United Nations. 2020 May 11. URL: http://digitallibrary.un.org/record/3863213 [accessed 2020-10-07]

Abbreviations

API: application programming interface
CNN: convolutional neural network
GCC: Gulf Corporation Council
JSON: JavaScript Object Notation
LDA: latent Dirichlet allocation
NMF: nonnegative matrix factorization

Edited by G Eysenbach, R Kukafka; submitted 17.07.20; peer-reviewed by R Sun, M El Tantawi; comments to author 25.09.20; revised version received 10.10.20; accepted 16.11.20; published 08.12.20.

Please cite as:
Alshalan R, Al-Khalifa H, Alsaeed D, Al-Baity H, Alshalan S
Detection of Hate Speech in COVID-19–Related Tweets in the Arab Region: Deep Learning and Topic Modeling Approach
J Med Internet Res 2020;22(12):e22609
URL: http://www.jmir.org/2020/12/e22609/
doi:10.2196/22609
PMID:33207310

Original Paper

Rapid Implementation of Video Visits in Neurology During COVID-19: Mixed Methods Evaluation

Erika A Saliba-Gustafsson[1*], BSc, MMSc, PhD; Rebecca Miller-Kuhlmann[2*], MD; Samantha M R Kling[1], RDN, PhD; Donn W Garvert[1], MS; Cati G Brown-Johnson[1], PhD; Anna Sophia Lestoquoy[1], MPH; Mae-Richelle Verano[1], BA; Laurice Yang[2], MD, MHA; Jessica Falco-Walter[2], MD; Jonathan G Shaw[1], MD, MS; Steven M Asch[1,3], MD, MPH; Carl A Gold[2], MD, MS; Marcy Winget[1], PhD

[1]Primary Care and Population Health, Stanford University School of Medicine, Stanford University, Palo Alto, CA, United States
[2]Department of Neurology & Neurological Sciences, Stanford University School of Medicine, Stanford University, Palo Alto, CA, United States
[3]Center for Innovation to Implementation (Ci2i), VA Palo Alto Health Care System, Menlo Park, CA, United States
[*]these authors contributed equally

Corresponding Author:
Erika A Saliba-Gustafsson, BSc, MMSc, PhD
Primary Care and Population Health
Stanford University School of Medicine
Stanford University
Stanford MSOB
1265 Welch Rd, Mail Code 5475
Palo Alto, CA, 94305
United States
Phone: 1 650 723 1146
Email: esalgus@stanford.edu

Abstract

Background: Telemedicine has been used for decades. Despite its many advantages, its uptake and rigorous evaluation of feasibility across neurology's ambulatory subspecialties has been sparse. However, the COVID-19 pandemic prompted health care systems worldwide to reconsider traditional health care delivery. To safeguard health care workers and patients, many health care systems quickly transitioned to telemedicine, including across neurology subspecialties, providing a new opportunity to evaluate this modality of care.

Objective: To evaluate the accelerated implementation of video visits in ambulatory neurology during the COVID-19 pandemic, we used mixed methods to assess adoption, acceptability, appropriateness, and perceptions of potential sustainability.

Methods: Video visits were launched rapidly in ambulatory neurology clinics of a large academic medical center. To assess adoption, we analyzed clinician-level scheduling data collected between March 22 and May 16, 2020. We assessed acceptability, appropriateness, and sustainability via a clinician survey (n=48) and semistructured interviews with providers (n=30) completed between March and May 2020.

Results: Video visits were adopted rapidly; overall, 65 (98%) clinicians integrated video visits into their workflow within the first 6 implementation weeks and 92% of all visits were conducted via video. Video visits were largely considered acceptable by clinicians, although various technological issues impacted their satisfaction. Video visits were reported to be more convenient for patients, families, and caregivers than in-person visits; however, access to technology, the patient's technological capacity, and language difficulties were considered barriers. Many clinicians expressed optimism about future utilization of video visits in neurology. They believed that video visits promote continuity of care and can be incorporated into their practice long-term, although several insisted that they can never replace the in-person examination.

Conclusions: Video visits are an important addition to clinical care in ambulatory neurology and are anticipated to remain a permanent supplement to in-person visits, promoting patient care continuity, and flexibility for patients and clinicians alike.

(J Med Internet Res 2020;22(12):e24328) doi:10.2196/24328

KEYWORDS

teleneurology; telemedicine; telehealth; ambulatory neurology; video visits; COVID-19; implementation; outcomes; video; neurology; mixed methods; acceptability; sustainability

Introduction

Telemedicine has been used for decades, yet its uptake and rigorous evaluation of feasibility across neurology's ambulatory subspecialties has been sparse [1,2]. Studies have shown several advantages of telemedicine including improved access to care [3,4], less travel burden, and fewer associated out-of-pocket and health care costs [1,5-8]. Nonetheless, integration of virtual visits into ambulatory neurology care, where the physical examination and nuanced communication play a strong role, has been received with some hesitation [7,9-13]. Further, issues related to technology, compensation, payor reimbursement, policy, hardware/software costs, credentialing, liability, and requirements for in-person evaluations prior to virtual care, have also limited its adoption [1,8,11,13-17].

The declaration of COVID-19 as a pandemic in March 2020 prompted health care systems worldwide to reconsider traditional health care delivery and quickly transition to telemedicine [18-21]. In regions with rapidly increasing COVID-19 cases and early stay-at-home directives, health care systems rapidly built and implemented infrastructure for telemedicine technologies to protect health care workers and patients, and conserve personal protective equipment. To further support this pivot and maintain health care access, the United States loosened previously stringent federal regulations on reimbursements, licensing, and Health Insurance Portability and Accountability Act (HIPAA) compliance [22,23]. In neurology specifically, the urgent need to provide care safely to patients with chronic illnesses while preventing disease transmission during clinic visits, led to rapid implementation of video visits across all subspecialties [24-29].

In this study, we evaluate the implementation of video visits in Stanford Health Care's (California, United States) ambulatory neurology clinics using mixed methods to assess adoption and explore clinicians' perspectives on the acceptability, appropriateness, and sustainability of this broad expansion.

Methods

Setting

This quality improvement project received a nonresearch determination by Stanford University's Institutional Review Board (IRB-55644). It was conducted at Stanford University's Department of Neurology and Neurological Sciences, which includes 11 ambulatory subspecialties staffed by 60 physicians and 8 advanced practice providers (APPs). As previously described [27], at the beginning of the local COVID-19 stay-at-home directive in March 2020, approximately 50 in-clinic computers were video visit–enabled, and 50 additional computer devices were readied for remote use by providers. Clinic staff converted >90% of scheduled in-person visits to video. In-person visits were reserved for procedures such as autonomic testing, implanted device programming, injections, electromyography, and electroencephalograms, as well as patients determined either during previsit screening or initial video visit to require a full examination.

Primary Outcomes, Data Collection, and Analysis

Overview

The primary implementation outcomes assessed were adoption, acceptability, appropriateness, and perceived sustainability of video visits [30], as described in Table 1. Clinician-level scheduling data were used to assess adoption. A combination of a clinician survey and interviews were used to assess acceptability, appropriateness, and perceived sustainability. Clinicians included physicians and APPs.

Table 1. Implementation outcomes, definitions, and data sources applied in the evaluation of the implementation of video visits in ambulatory neurology.

Proctor et al's [30] implementation outcomes	Definitions	Data sources
Adoption	Uptake of video visits in ambulatory neurology	Scheduling data
Acceptability	Clinicians' overall satisfaction with video visits	Clinician interviews and clinician survey
Appropriateness	"Fit for purpose": clinicians' perceived suitability and practicability of video visits for a successful patient visit to achieve similar patient outcomes to an in-person visit	Clinician interviews and clinician survey
Sustainability	Clinicians' views on the future use of video visits in their practice	Clinician interviews and clinician survey

Scheduling Data

Clinician-level scheduling data were extracted for all clinicians who had the opportunity to conduct video visits. The first week of the stay-at-home order (March 15-21, 2020) was considered a transition week where the physician champion onboarded physicians onto the video visit platform. To assess our primary outcome, adoption, scheduling data from the 8-week implementation period (March 22 through May 16, 2020) was used for analysis; the transition week was excluded. Adoption was assessed in three ways: (1) proportion of clinicians who conducted video visits, (2) proportion of visits completed via video during the COVID-19 pandemic, and (3) proportion of all "expected" visits that were done via video, where "expected" was defined as the estimated number of visits completed had COVID-19 not occurred (ie, visit volume from the same calendar period in the prior year, March 24 through May 18, 2019).

The number of visits completed during the implementation period (ie, numerator) and comparator period (ie, denominator) included visit types that could be feasibly conducted in-person or via video. Visit types inherently requiring in-person patient interaction, such as procedures and research visits, were excluded. The percentage of "expected" visits is shown for video and in-person visits separately, along with the percentage of lost potential visits. The proportion of all "expected" visits completed is a measure of the proportion of "saved" or "protected" visits attributable to video visits during the COVID-19 stay-at-home period.

Clinician Survey

We developed a 20-item survey informed by early interview findings and reflections from the clinical improvement team. The survey was administered via Stanford's REDCap platform [31] and emailed to all clinicians using video visits in May 2020, followed by two reminders. To increase response rates, clinicians were encouraged to complete the survey during division meetings and reminded personally by clinical leaders. Descriptive analyses were conducted on 5 items (Multimedia Appendix 1): (1) need for and timing of in-person visits to supplement video visits, (2) top three concerns regarding video visits, (3) top three reasons for excitement for video visits, (4) video visits support of overall clinician wellbeing, and (5) video visit usage to shift uncompensated to compensated work. The number and percentage of clinicians indicating each response are reported for questions 1-3 and for those indicating "agree" or "strongly agree" for questions 4 and 5.

All quantitative data were processed and analyzed using SAS (Version 9.4; SAS Institute Inc), R (Version 4.0.2; R Foundation for Statistical Computing) [32] and associated packages [33-36].

Clinician Interviews

We designed the interview guide (Multimedia Appendix 2) to encompass three implementation outcomes of interest: acceptability, appropriateness, and perceived sustainability [30]. A purposive sample of 47 clinicians who had conducted at least one video visit were invited to participate in semistructured phone interviews held between March and May 2020. Stakeholders were intentionally recruited from different neurology subspecialties to obtain a representative departmental sample. In total, 30 clinicians were interviewed and interviews lasted 25 minutes, on average.

Interview notes and transcripts were used for analysis. To ensure anonymity, identifiable information was removed from transcripts and subspecialties with <5 clinicians were grouped. Data were analyzed using both deductive and inductive approaches. Deductive codes were derived from the implementation outcomes of interest [30]. During analysis, barriers, facilitators, and other emergent themes were identified and coded. A multiphase analysis approach leveraged rapid analytic procedures (eg, template summaries) to extract early themes, consensus coding of transcript summaries to produce interim results, and a matrix analysis [37] of interview excerpts for final comparison of themes across neurology subspecialties. First, individual interview transcripts were summarized independently by two coauthors into a templated summary document. After review and consensus discussion, transcript summaries were consolidated into a matrix to identify themes and allow for comparison across participants using Microsoft Excel (Microsoft Corp). Four qualitative coauthors met weekly for two months to discuss preliminary results and achieve reporting consensus.

Mixed Methods Analysis

Quantitative and qualitative data were consolidated during analysis and interpreted in parallel to understand the impact of video visits in ambulatory neurology more comprehensively. This approach allowed us to harness strengths and offset weaknesses of the two methodologies [38,39]. We were also able to identify converging and diverging issues regarding adoption, acceptability, appropriateness, and perceived sustainability of video visits.

Results

Participants

Out of the 68 clinicians in the department's 11 ambulatory subspecialties, 66 clinicians conducted regular video visits during the 8-week implementation period and were included in analyses (two clinicians were on leave and thus excluded). Table 2 summarizes the number of surveyed and interviewed clinicians, by subspecialty. In total, 48 (73%) clinicians responded to the survey and 30 (45%) were interviewed. In total, 21 (32%) clinicians participated in both the survey and interviews. Only 9 (14%) did not participate in either modality.

Table 2. Number of ambulatory neurology clinicians who completed the video visit survey and were interviewed, by subspecialty.

Subspecialty	Interview respondents, n	Survey respondents, n	Total clinicians, n
Autonomic, neuro-oncology, and neuro-ophthalmology[a]	3	4	9
Epilepsy	4	9	10
General neurology	2	5	5
Headache	4	3	5
Memory	3	4	6
Movement disorders	3	8	9
Neuroimmunology	3	5	5
Neuromuscular	4	5	8
Stroke	4	5	9
Total	30 (45%)[b]	48 (73%)[c]	66 (100%)

[a] The autonomic, neuro-oncology, and neuro-ophthalmology subspecialties had <5 clinicians and were therefore grouped to ensure anonymity.

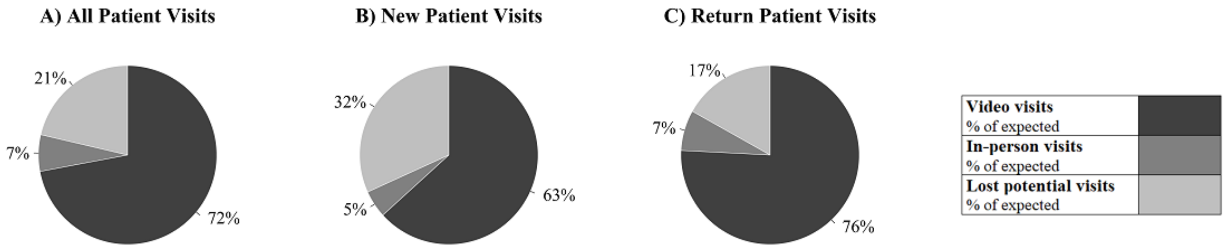
[b] Of the 30 interviewees, 29 were physicians and 1 was an advanced practice provider.

[c] Of the 48 survey respondents, 41 were physicians and 7 were advanced practice providers.

Adoption

During the transition week, the clinician champion onboarded 30 of the 66 physicians included in this study, and 129 video visits were conducted. Prior to this transition week, clinicians in these ambulatory neurology clinics had not used video visits. During the 8-week implementation period, video visit adoption was high based on both percentage of clinicians using them, and percentage of visits completed via video. Within the first two weeks, 52 (79%) clinicians integrated video into their practice, increasing to 65 (98%) clinicians by week 6. Almost all (92%) visits were conducted via video and adoption of video visits was high for both new and return patient visits (93% and 91% of completed visits, respectively). The total number of completed visits, conservatively estimated, was almost 80% of expected visits based on the 2019 comparison period (Figure 1A). Clinics recuperated more expected return patient (83%) than new patient visit volumes (68%; Figures 1B-C).

Figure 1. Percentage of video visits conducted between March 22 and May 16, 2020, as a proportion of (A) all "expected" visits overall, (B) "expected" new patient visits, and (C) "expected" return patient visits. The "expected" number of visits is defined as the number of completed visits in the comparable 2019 time period. Data only reflect operational scheduling data captured through the EPIC Hyperspace Platform. The data do not reflect no shows, visits conducted on other HIPAA-compliant software or visits that reverted to phone calls. (HIPAA: Health Insurance Portability and Accountability Act).



Acceptability

Overall acceptability of video visits was high, but clinicians experienced technological and scheduling issues, and identified additional needs. Two themes emerged related to acceptance: technology and workflow efficiency.

Technology

While some clinicians experienced a smooth transition to video visits and seamless connection with patients, others dealt with numerous technical issues, jeopardizing their overall satisfaction. Indeed, technology was one of the top 3 concerns regarding video visits for 30 (63%) survey respondents (Table 3).

Table 3. Ambulatory neurology clinicians' (n=48) top 3 areas of concern related to video visits as reported through survey data.

With regard to video visits, clinicians (n=48) were concerned about... (indicated top 3)	Survey respondents, n (%)
Technological limitations	30 (63)
Being able to engage in training and education of residents and fellows	21 (44)
Missing/losing the in-person connection/relationship with patients	18 (38)
Including interpreters on video calls	17 (35)
Difficulties arranging and completing necessary follow-ups after the video visit	14 (29)
Insurance reimbursements for video visits are not the same as for in-person visits	8 (17)
Patients' unwillingness to come into clinic for requested in-person visits in the future	7 (15)
Press-Ganey scores	5 (10)
Patient expectations to have video visits as an option	3 (6)
Maintaining access to readily available technology and equipment needed for video visits	2 (4)
Other (providers listed various reasons)	9 (19)

Several video functionalities were suggested for a successful visit, including the following: (1) *screen sharing* to facilitate patient education and explain imaging results, (2) a *waiting room function* to replicate "stepping out of the room" when engaging with trainees, (3) a *chat box* for troubleshooting, (4) *file sharing* capabilities, (5) *screenshot* capabilities to support efficient charting, and (6) *multiperson teleconferencing* to include other members of the multidisciplinary team, interpreters, trainees, and family members in different physical locations. These functionalities were not available during the study period.

> *...when we're discussing end-of-life care issues, to have the physician and respiratory therapist, social worker and palliative care physician, all present in the same 'virtual room', that would be really nice.* [MD19]

Engaging in education of trainees was one of the top 3 concerns for 21 (44%) surveyed clinicians (Table 3). The initial interface available through the electronic medical record (EMR) only allowed for a two-way call; therefore, parallel software was needed to include trainees.

Workflow Efficiency

Perceptions of workflow efficiency were mixed. Major issues included rigid video visit scheduling, note-taking efficiency, and previsit planning. Several clinicians mentioned that, provided there were no technological issues, video visits helped them stay on schedule. However, hard time limits set by the video visit platform caused frustration. Clinicians could not notify the next patient when running late nor initiate the next visit earlier than scheduled as they might do at the clinic.

> *…it's much easier for someone to wait in a physical waiting room if you do happen to run over with a patient who's before them. It's a lot harder or less acceptable if someone is waiting to see you via video and staring at their computer screen for 10 or 15 minutes. … You can't open up another video and say, 'I'm here. I'm just with somebody else, and I'm going to switch back.'* [MD10]

Views on documentation during a video call also differed. Note-taking was considered easier by some during video visits, reducing after-hours charting time. Others were unable to document simultaneously, making them less efficient. Dictation was considered a possible solution.

> *…unless you get the screen set up correct[ly], it actually can be really hard to document and type while doing the video visit. The typing can be noisy on the patient's end…* [MD5]

However, most surveyed clinicians (n=34; 71%) agreed or strongly agreed that video visits allowed them to shift uncompensated to compensated work (ie, scheduling a video visit to address concerns that would previously have been managed through either EMR messaging or unscheduled and uncompensated phone calls).

The lack of integration of medical assistants (MAs) in the video visit workflow, resulting in a lack of previsit charting and medication reconciliation, was another concern. Incorporating MAs, or even involving the patient in their own previsit preparation, was deemed necessary for video visits to be sustained.

> *I'd like to see it be more efficient… In advance of the visit, it would be nice if they [patients] could do their medication reconciliation. …enter in the last set of vitals that they took for themselves. …enter in their own review of symptoms and chief complaints.* [MD16]

Appropriateness: Suitability, Usefulness, and Practicability of Video Visits

Three themes emerged related to appropriateness of video visits: benefits and barriers, physical examination, and continuity of care, each of which are described below.

Benefits and Barriers of Video Visits for Patients and Families/Caregivers

Clinicians highlighted several benefits of video visits for patients and families/caregivers, including convenience, impact on travel and cost, and seeing the patient's home environment. Barriers noted included access to technology, patient’s technological capacity, and language. Video visits were considered advantageous and convenient, saving patients time and money, particularly for older adults and those who travel long distances for appointments. Survey respondents (n=37; 77%) agreed that saving patients unnecessary travel was one of the top 3 benefits (Table 4). Several clinicians supported referring patients to local laboratories to further avoid unnecessary travel. Clinicians also noted that for patients who required assistance (eg, patients with dementia, epilepsy, or mobility issues), the video visit alleviated the travel burden on families/caregivers.

> *…it's very convenient for our patients, especially those who are elderly or have neurologic issues. [It] Spares them unnecessary travel and costs...* [MD16]

Table 4. Ambulatory neurology clinicians’ (n=48) top 3 areas of excitement related to video visits as reported through survey data.

With regard to video visits, clinicians (n=48) were excited about... (indicated top 3)	Survey respondents, n (%)
Saving patients from unnecessary travel	37 (77)
Increased access for vulnerable populations	33 (69)
Ability to see my patients from my home or nonclinic location	23 (48)
Reduced uncompensated work	18 (38)
Flexible scheduling of patient visits	17 (35)
Ability to see patients in their home environment	3 (6)
Ability to connect with patients’ caregivers/family members	3 (6)
Other (providers listed various reasons)	4 (8)

Several clinicians saw value in seeing patients in their home. Not only did it allow more family involvement, but clinicians could also troubleshoot daily functioning issues or modify the patient’s environment by directly assessing fall hazards, medications, and home devices.

> *…maybe you see … a lot of stuff strewn over the floor and that is why they're falling so much … you might observe things … that might impact their neurological disease on a day to day basis … or you don't think to ask about … in a clinic visit because you're not seeing the world that they're actually living in.* [MD25]

Even though increased access to vulnerable populations was rated as one of the top 3 benefits of video visits (n=33; 69%; Table 4), numerous clinicians mentioned that some patients, particularly older adults and lower-income patient populations (eg, unhoused individuals or rural farm workers), lacked the necessary access to technology and technological capability to support a video call. In the absence of a supportive family member/caregiver, video visits with patients with cognitive, hearing, or visual impairment were also considered nonideal.

> *... not suited are patients with some mild to moderate cognitive impairment. ...they have more trouble interacting with the physician and understanding who they’re talking to.* [MD25]

Language barriers were also considered a possible limitation of video visit utilization by both interviewees and survey respondents. Specifically, 17 (35%) surveyed clinicians indicated inclusion of interpreters was one of their top 3 concerns (Table 3).

Virtual Physical Examinations

Clinicians reported that video visits were superior to a phone call, allowing them to gather more information than just a medical history. Although exam needs varied by clinician and subspecialty, several clinicians described that despite challenges, they were pleasantly surprised to be able to perform several modified physical examinations over video. Nevertheless, many stated that the inability to perform a hands-on physical examination was a limitation of video visits that was best paired with a timely in-person follow-up. Several clinicians mentioned that virtual exams were more time-consuming and occasionally required assistance from a caregiver to position the phone to properly observe the patient, perform certain physician-directed exams, confirm what the patient says, or catch the patient if s/he is at risk of falling.

> *You have to have a caregiver, you can’t check gait without it [a caregiver] or the patient needs some sort of stand or desk or something where they can put the computer down at the end of a hallway and then asking the patient to find a hallway that they can walk back and forth is really the best way to do it. But a caregiver probably works best.* [MD21]

Additional limitations were related to the patient’s immediate environment, including adequate space for the patient to move around, and sufficient lighting. Camera positioning was also critical. If positioned too close, the clinician lacked sufficient visual field to observe the patient’s entire body. Clinicians also noted that occasionally patients took their video call at inappropriate times (eg, while working or driving), despite previsit counseling.

Promoting Continuity of Care

Clinicians considered video visits beneficial to ensure continuity of care for chronic conditions. There was general agreement that video visits are best suited for established patients, especially those who are stable/uncomplicated, or for a quick checkup without extensive examination or testing. Although several clinicians felt that most patients were appropriate for video visits, the majority agreed that new patient visits, and patients with acute conditions and declining health, were less suited due to relative ease of the complete physical examination when conducted in person. Many preferred to first see the patient in person before determining whether further follow-up care can be provided, at least partly, over video. Clinicians also recognized that some patient populations encounter significant barriers to attending the clinic in person and that video offers an opportunity to continue care.

> *We deal with progressive conditions, and once you get into that moderate to advanced stage of illness it just becomes not worth their effort to come anymore. It's too difficult on the patient, it's too difficult on the family, and so we lose contact with the vast majority of our patients as they become more advanced. ...telemedicine actually might allow continuity of care into the more advanced stages…* [MD12]

Perceived Sustainability

Most clinicians were positive toward video visits and believed they could incorporate video into their practice long-term, although several insisted that video cannot replace a full in-person examination. Moreover, 40 (83%) surveyed clinicians agreed or strongly agreed that video visits supported their overall well-being. The general view was that for long-term sustainability, patient video visits will need to be selected carefully to optimize care and respect preferences. Patients' suitability for video visits would need to be determined during scheduling based on several criteria (eg, physical examination needs, patient's technological capacity and demographics, new versus return). However, previsit screening and triage to determine which patients are best suited for video was considered burdensome, which could compromise their sustainability. Seldom was the video visit itself considered a good tool for triage.

> *I would like to be able to offer it for my patients who are a little less mobile and who come from far away. I don't think it will completely replace in-person visits, but if we can alternate visits and do in-person versus telehealth, I think that'll be huge.* [MD21]

Most survey respondents (n=39; 81%) agreed that video visits should be supplemented with in-person visits. The recommended frequency of supplemental in-person visits varied among the respondents: 6 (15%) recommended quarterly, 11 (28%) biannually, 17 (44%) annually, and 5 (13%) every 2 years. A concern mentioned by a minority of interviewees and 7 (15%) surveyed clinicians was that patients may find video too convenient and opt out of recommended in-person visits. In contrast, 9 (19%) survey respondents with representation from epilepsy, memory, headache, stroke, and neuro-oncology, indicated that an all-video practice would be feasible for their patient population.

Discussion

Principal Findings

Video visits were rolled out rapidly at Stanford University's Department of Neurology and Neurological Sciences during the COVID-19 pandemic. A high proportion of visits were conducted via video at volumes near prepandemic volumes, indicating robust and rapid adoption by both clinicians and patients. This necessary and widespread adoption allowed for thorough assessment of new opportunities and barriers in video-based care across almost all ambulatory neurology subspecialties.

Clinicians recognized the patient benefits of video visits, including saving patients unnecessary travel and increasing access to vulnerable populations, as is well-documented [1,3-8]. As patients with chronic progressive neurologic conditions lose function, in-person visits become increasingly challenging and care by video is a welcome new supplemental resource. Video visits could improve the experience for these patients and others, but various challenges remain. As clinicians noted, the most vulnerable high-need patients were often the hardest to care for virtually. Vulnerability is not uniform across patients and could be related to diagnosis (eg, patients with sensory or cognitive deficits), support (eg, caregiver status), socioeconomic status, or access (eg, poor access to technology). New creative solutions are being explored to address these challenges at various levels, with the goal of rendering virtual care an effective catchment for all. For instance, a pilot program delivering tablets to patients cared for by the US Department of Veterans Affairs has been successful in addressing concerns related to diagnosis and technology [40]. Further, municipality-level internet and broadband for all could address access issues [41].

Clinicians reported that video visits were beneficial for seeing the patient's home and meeting caregivers and family members. Observing patients at home is a documented benefit of video visits; it can reduce patient and caregiver anxiety, and allow assessment of fall hazards and habitual behaviors that are not always possible to capture during an in-person visit [6,42]. These benefits were rarely identified within the top 3 benefits of video visits in the survey, which is likely related to the many perceived benefits of video visits as well as the diverse needs of neurology subspecialties' patient populations. Interviewed clinicians emphasized that the presence of a caregiver is often essential to a successful video visit. For example, a caregiver is essential for collateral history when the patient has cognitive impairment or for safely examining gait for a complaint of parkinsonism, but not for a younger patient presenting with headache. Communicating the importance of including caregivers in certain video visits is thus important for optimal care provision.

Previously, the perceived limitations of virtual physical examinations have been a source of hesitant incorporation of telemedicine into neurology practice [10-12]; however, virtual physical examinations are possible, especially to supplement heavily history-based presentations [1,6,43]. A previous investigation demonstrated that virtual exams were adequate even for many patients with multiple sclerosis and neuroimmunology issues [44]. Presence of a caregiver can further facilitate a detailed and complete medical history and examination [28]. In this study, several clinicians reported that video cannot replace a full in-person examination, but others reported that the virtual physical examination was more feasible than expected. Best practices that were followed across clinics included consenting patients and explaining video limitations prior to the visit, and scheduling patients for in-person follow-up as soon as they are able to do so safely whenever an in-person examination was deemed important during the video visit. As clinics adjust to new routines and workflows, measures are also being taken to further optimize the virtual exam experience. For instance, the department has organized teleneurology professional development webinars on optimal virtual physical examination techniques [23,45], and guidance documents on this topic have also been developed internally. Together with time and experience, these efforts may further increase clinician comfort with the virtual physical examination.

Clinicians were keen about the ability to work remotely, reduction in uncompensated work, and flexible scheduling facilitated by video visit use. In the current configuration, video visits are not constrained by clinic spaces nor specific staffing needs, thereby allowing clinicians flexibility to vary timing and length of both their overall clinic session and individual patient appointments. This is a clear benefit, as lack of flexibility in work is a key driver of burnout [46], and even more salient given increased family and childcare duties for many due to the pandemic. Furthermore, reducing uncompensated work has potential well-being benefits. Historically, clinicians with full clinic schedules answered patient messages and phone calls uncompensated outside of normal clinical hours; this is another driver of physician burnout [47]. Video visits have relatively fewer constraints than in-person visits that are often booked out for months, and could enable the ability to address a new patient concern in short order “on the clock” and compensated. Further, most respondents also reported that video visits support their well-being, which is key in a specialty where an estimated 60% of clinicians reported at least one symptom of burnout even before the pandemic [48].

Perceived sustainability of video visits was high; clinicians emphasized a desire for video visits to remain a permanent fixture in their practice. However, the proportion of video visits and in-person visits within their practice was anticipated to change in their postpandemic practice and vary by individual providers and subspecialty. Most clinicians preferred to have a practice that uses both video and in-person visits, but more than half (57%) expressed a preference for less frequent in-person visits (ie, every 1 or 2 years). Perhaps most surprisingly, nearly 20% of clinicians from a variety of neurology subspecialties indicated that an all-virtual practice was possible. This variation in preferences may reflect the wide variation in the nature and examination demands of neurologic subspecialties, with subspecialties known to be especially exam-dependent (eg, movement and neuromuscular disorders) notably absent from those indicating feasibility of an entirely virtual practice.

The capabilities of mobile technology and remote monitoring have facilitated great advances in telemedicine [2,6,8,14]; however, several challenges remain and need to be addressed to enable long-term sustainability of video visits. Audio and video connectivity issues may be partially ameliorated by software updates and through increased experience with telemedicine, particularly for clinicians. On the other hand, some technological issues experienced by patients could potentially be addressed with targeted efforts, including more previsit planning and education with a staff member. To enhance the video visit experience for both patients and clinicians, pertinent features included virtual waiting rooms and multiuser interfaces to promote incorporation of trainees, interpretation services, and multidisciplinary care. Some of these features were added in subsequent updates and parallel workflows were developed to use other HIPAA-compliant platforms for visits including trainees or interpreters. In our experience, this parallel platform and approach were important for resolving some issues of functionality and service; however, the lack of integration and interoperability with standard EMR video functionality could impact sustainability.

Further analysis of the workflow around telemedicine in ambulatory neurology is needed as there is growing recognition that workflow is nontrivial and complex [49]. In this study, initial workflows for video visits did not include MAs taking patient vitals and performing medication reconciliation, key aspects of in-person visits for new patients. Clinic processes should reincorporate MAs in video previsits, perhaps through phone or video visits in the days prior to the clinical video visit. As clinical workflows evolve to sustain video visit use and meet the needs of the most vulnerable patient populations, these video previsits could potentially also provide patient education on optimal lighting, how to position devices, space, and presence of caregivers.

Strengths and Limitations

Perceptions of clinicians using video visits were captured in real time during early rapid implementation through interviews and/or a survey across many ambulatory neurology subspecialties. However, given its nascent state, operational scheduling data only captured data of video visits conducted via the EPIC Hyperspace Platform and did not consistently reflect visits conducted on other HIPAA-compliant software, such as Zoom. Therefore, the data presented herein likely underestimate the actual total number of completed visits. Further, when video visits were diverted to phone calls or other software during the 8-week implementation period, the scheduling system incorrectly categorized the encounter as cancelled, no show, or patient left without being seen. This limitation has been recently addressed, presenting an opportunity to investigate the impact of video visits on clinic utilization in future investigations. Finally, views of other essential health care staff, such as patient care coordinators and MAs, who performed essential activities (eg, scheduling and confirming patients’ technology capabilities), as well as residents and patients/caregivers, are not presented here.

Conclusions

Video visit adoption was rapid at Stanford’s ambulatory neurology clinics. Almost all clinicians conducted video visits by the eighth week of implementation and achieved near-normal patient volumes during the COVID-19 pandemic stay-at-home order. Despite the sudden change in workflow, clinicians largely expressed positive views toward video visits; clinicians supported permanent integration of video visits and noted them to be conducive to physician well-being. However, overall clinician satisfaction was impacted by technological issues, limitations with the physical examination, and challenges accessing vulnerable patient populations. Although our mixed methods evaluation confirmed the success of video visits in all subspecialty neurology clinics across many dimensions, innovations must be developed to address their limitations. Additional solutions are also needed for the most vulnerable patient populations. A crucial next step for optimization is to understand patients' experiences and preferences.

Acknowledgments

We would like to extend our gratitude to all participating clinicians who shared their insights with us. Thank you to Dr Frank Longo, Dr Yuen So, Alison Kerr, Sophia Loo, and Dana Gonzales for their support of this clinical innovation. We also thank Dr David Larson and the leadership of the Improvement Capability Development Program for arranging the partnership between the Department of Neurology & Neurological Sciences and the Evaluation Sciences Unit.

This project was supported by Stanford Health Care as part of the Improvement Capability Development Program. The Stanford REDCap platform was used for quantitative survey data collection. The Stanford REDCap platform [31] is developed and operated by the Stanford Medicine Research information technology team. The REDCap platform services at Stanford are subsidized by both the Stanford School of Medicine Research Office and the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through grant UL1 TR001085.

Authors' Contributions

RMK, SMRK, CGBJ, LY, JFW, JGS, SMA, CAG, and MW contributed to conception and design. RMK, LY, JFW, and CAG contributed to implementation. EASG, SMRK, and DWG contributed to data acquisition, analysis, and interpretation. CGBJ, ASL, and MRV contributed to data acquisition and analysis. EASG, RMK, SMRK, MW contributed to manuscript drafts and significant revisions. CGBJ, DWG, ASK, MRV, LY, JFW, JGS, SMA, and CAG contributed to manuscript revisions. All authors gave final approval of the published version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Selected questions from 20-item neurology video visit clinician survey administered electronically via the REDCap survey tool.

[DOCX File , 22 KB - jmir_v22i12e24328_app1.docx]

Multimedia Appendix 2

Interview guide used to understand ambulatory neurology clinicians' views on the acceptability, appropriateness, and sustainability of video visits in their practice.

[DOCX File , 15 KB - jmir_v22i12e24328_app2.docx]

References

1. Hatcher-Martin JM, Adams JL, Anderson ER, Bove R, Burrus TM, Chehrenama M, et al. Telemedicine in neurology. Neurology 2020 Jan 07;94(1):30-38. [doi: 10.1212/wnl.0000000000008708]
2. Dorsey ER, Glidden AM, Holloway MR, Birbeck GL, Schwamm LH. Teleneurology and mobile technologies: the future of neurological care. Nat Rev Neurol 2018 May 6;14(5):285-297. [doi: 10.1038/nrneurol.2018.31] [Medline: 29623949]
3. Davis LE, Harnar J, LaChey-Barbee LA, Pirio Richardson S, Fraser A, King MK. Using teleneurology to deliver chronic neurologic care to rural veterans: analysis of the first 1,100 patient visits. Telemed J E Health 2019 Apr;25(4):274-278. [doi: 10.1089/tmj.2018.0067] [Medline: 30016207]
4. Demaerschalk BM, Berg J, Chong BW, Gross H, Nystrom K, Adeoye O, et al. American telemedicine association: telestroke guidelines. Telemed J E Health 2017 May;23(5):376-389 [FREE Full text] [doi: 10.1089/tmj.2017.0006] [Medline: 28384077]
5. Appireddy R, Khan S, Leaver C, Martin C, Jin A, Durafourt BA, et al. Home virtual visits for outpatient follow-up stroke care: cross-sectional study. J Med Internet Res 2019 Oct 07;21(10):e13734 [FREE Full text] [doi: 10.2196/13734] [Medline: 31593536]
6. Ben-Pazi H, Browne P, Chan P, Cubo E, Guttman M, Hassan A, International Parkinson and Movement Disorder Society Telemedicine Task Force. The promise of telemedicine for movement disorders: an interdisciplinary approach. Curr Neurol Neurosci Rep 2018 Apr 13;18(5):26. [doi: 10.1007/s11910-018-0834-6] [Medline: 29654523]
7. Mammen JR, Elson MJ, Java JJ, Beck CA, Beran DB, Biglan KM, et al. Patient and physician perceptions of virtual visits for Parkinson's disease: a qualitative study. Telemed J E Health 2018 Apr;24(4):255-267. [doi: 10.1089/tmj.2017.0119] [Medline: 28787250]
8. Schneider RB, Biglan KM. The promise of telemedicine for chronic neurological disorders: the example of Parkinson's disease. Lancet Neurol 2017 Jul;16(7):541-551. [doi: 10.1016/S1474-4422(17)30167-9] [Medline: 28566190]

9. Howard IM, Kaufman MS. Telehealth applications for outpatients with neuromuscular or musculoskeletal disorders. Muscle Nerve 2018 Oct 17;58(4):475-485. [doi: 10.1002/mus.26115] [Medline: 29510449]
10. Helleman J, Kruitwagen ET, van den Berg LH, Visser-Meily JMA, Beelen A. The current use of telehealth in ALS care and the barriers to and facilitators of implementation: a systematic review. Amyotroph Lateral Scler Frontotemporal Degener 2020 May 26;21(3-4):167-182. [doi: 10.1080/21678421.2019.1706581] [Medline: 31878794]
11. Hassan A, Dorsey ER, Goetz CG, Bloem BR, Guttman M, Tanner CM, et al. Telemedicine use for movement disorders: a global survey. Telemed J E Health 2018 Dec;24(12):979-992. [doi: 10.1089/tmj.2017.0295] [Medline: 29565764]
12. Geronimo A, Wright C, Morris A, Walsh S, Snyder B, Simmons Z. Incorporation of telehealth into a multidisciplinary ALS clinic: feasibility and acceptability. Amyotroph Lateral Scler Frontotemporal Degener 2017 Nov 05;18(7-8):555-561. [doi: 10.1080/21678421.2017.1338298] [Medline: 28678542]
13. George BP, Scoglio NJ, Reminick JI, Rajan B, Beck CA, Seidmann A, et al. Telemedicine in leading US neurology departments. Neurohospitalist 2012 Oct 17;2(4):123-128 [FREE Full text] [doi: 10.1177/1941874412450716] [Medline: 23983876]
14. Achey M, Aldred JL, Aljehani N, Bloem BR, Biglan KM, Chan P, International Parkinson and Movement Disorder Society Telemedicine Task Force. The past, present, and future of telemedicine for Parkinson's disease. Mov Disord 2014 Jun 17;29(7):871-883. [doi: 10.1002/mds.25903] [Medline: 24838316]
15. Wechsler LR. Advantages and limitations of teleneurology. JAMA Neurol 2015 Mar 01;72(3):349-354. [doi: 10.1001/jamaneurol.2014.3844] [Medline: 25580942]
16. Reider-Demer MM, Eliashiv D. Expanding the use of telemedicine in neurology: a pilot study. JMTM 2018 Oct;7(2):46-50. [doi: 10.7309/jmtm.7.2.6]
17. Patel UK, Malik P, DeMasi M, Lunagariya A, Jani VB. Multidisciplinary approach and outcomes of tele-neurology: a review. Cureus 2019 Apr 08;11(4):e4410 [FREE Full text] [doi: 10.7759/cureus.4410] [Medline: 31205830]
18. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 pandemic, lessons for the future. Telemed J E Health 2020 May 08;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]
19. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
20. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
21. Myers US, Birks A, Grubaugh AL, Axon RN. Flattening the curve by getting ahead of it: how the VA healthcare system is leveraging telehealth to provide continued access to care for rural veterans. J Rural Health 2020 Apr 13:1 [FREE Full text] [doi: 10.1111/jrh.12449] [Medline: 32282955]
22. Centers for Medicare and Medicaid Services. Medicare telemedicine health care provider fact sheet. 2020. URL: https://tinyurl.com/y4qz4am4 [accessed 2020-08-17]
23. American Academy of Neurology. Telemedicine and COVID-19 implementation guide. 2020. URL: https://tinyurl.com/y37xxkrv [accessed 2020-08-17]
24. Waldman G, Mayeux R, Claassen J, Agarwal S, Willey J, Anderson E, et al. Preparing a neurology department for SARS-CoV-2 (COVID-19). Neurology 2020 Apr 06;94(20):886-891. [doi: 10.1212/wnl.0000000000009519]
25. Guidon AC, Amato AA. COVID-19 and neuromuscular disorders. Neurology 2020 Apr 13;94(22):959-969. [doi: 10.1212/wnl.0000000000009566]
26. Klein BC, Busis NA. COVID-19 is catalyzing the adoption of teleneurology. Neurology 2020 Apr 01;94(21):903-904. [doi: 10.1212/wnl.0000000000009494]
27. Yang L, Brown-Johnson CG, Miller-Kuhlmann R, Kling SM, Saliba-Gustafsson EA, Shaw JG, et al. Accelerated launch of video visits in ambulatory neurology during COVID-19. Neurology 2020 Jul 01;95(7):305-311. [doi: 10.1212/wnl.0000000000010015]
28. Grossman SN, Han SC, Balcer LJ, Kurzweil A, Weinberg H, Galetta SL, et al. Rapid implementation of virtual neurology in response to the COVID-19 pandemic. Neurology 2020 May 01;94(24):1077-1087. [doi: 10.1212/wnl.0000000000009677]
29. Strowd RE, Strauss L, Graham R, Dodenhoff K, Schreiber A, Thomson S, et al. Rapid implementation of outpatient teleneurology in rural Appalachia: barriers and disparities. Neurol Clin Pract 2020 Jul 13:10.1212/CPJ.0000000000000906. [doi: 10.1212/cpj.0000000000000906]
30. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar 19;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
31. REDCap. Welcome to REDCap @Stanford. 2020. URL: https://redcap.stanford.edu/ [accessed 2020-08-17]
32. R Foundation for Statistical Computing. R: A language and environment for statistical computing. Vienna, Austria; 2020. URL: https://www.r-project.org/ [accessed 2020-08-17]
33. Auguie B, Antonov A. gridExtra: miscellaneous functions for Grid graphics. 2017. URL: https://CRAN.R-project.org/package=gridExtra [accessed 2020-08-17]

34. Wickham H, Pedersen TL. gtable: arrange Grobs in tables. RStudio. 2019. URL: https://CRAN.R-project.org/package=gtable [accessed 2020-08-17]
35. Urbanek S, Horner J. Cairo: R graphics device using Cairo graphics library for creating high-quality bitmap (PNG, JPEG, TIFF), vector (PDF, SVG, PostScript) and display (X11 and Win32) output. RStudio. 2020. URL: https://CRAN.R-project.org/package=Cairo [accessed 2020-08-17]
36. Barthelme S, Tschumperle D, Wijffels J, Assemlal HE, Ochi S. Imager: image processing library based on CImg. RStudio. 2019 Jun. URL: https://CRAN.R-project.org/package=imager [accessed 2020-08-17]
37. Averill JB. Matrix analysis as a complementary analytic strategy in qualitative inquiry. Qual Health Res 2002 Jul 01;12(6):855-866. [doi: 10.1177/104973230201200611]
38. Tariq S, Woodman J. Using mixed methods in health research. JRSM Short Rep 2013 Jun 31;4(6):2042533313479197 [FREE Full text] [doi: 10.1177/2042533313479197] [Medline: 23885291]
39. Bekhet AK, Zauszniewski JA. Methodological triangulation: an approach to understanding data. Nurse Res 2012 Nov 22;20(2):40-43. [doi: 10.7748/nr2012.11.20.2.40.c9442] [Medline: 23316537]
40. Slightam C, Gregory AJ, Hu J, Jacobs J, Gurmessa T, Kimerling R, et al. Patient perceptions of video visits using Veterans Affairs telehealth tablets: survey study. J Med Internet Res 2020 Apr 15;22(4):e15682 [FREE Full text] [doi: 10.2196/15682] [Medline: 32293573]
41. Benda NC, Veinot TC, Sieck CJ, Ancker JS. Broadband internet access is a social determinant of health!. Am J Public Health 2020 Aug;110(8):1123-1125. [doi: 10.2105/ajph.2020.305784]
42. Dorsey ER, Vlaanderen FP, Engelen LJ, Kieburtz K, Zhu W, Biglan KM, et al. Moving Parkinson care to the home.. Mov Disord 2016 Sep;31(9):1258-1262 [FREE Full text] [doi: 10.1002/mds.26744] [Medline: 27501323]
43. Blue R, Yang AI, Zhou C, De Ravin E, Teng CW, Arguelles GR, et al. Telemedicine in the era of coronavirus disease 2019 (COVID-19): a neurosurgical perspective. World Neurosurg 2020 Jul;139:549-557 [FREE Full text] [doi: 10.1016/j.wneu.2020.05.066] [Medline: 32426065]
44. Bove R, Garcha P, Bevan CJ, Crabtree-Hartman E, Green AJ, Gelfand JM. Clinic to in-home telemedicine reduces barriers to care for patients with MS or other neuroimmunologic conditions. Neurol Neuroimmunol Neuroinflamm 2018 Nov;5(6):e505 [FREE Full text] [doi: 10.1212/NXI.0000000000000505] [Medline: 30775409]
45. Boes CJ, Leep Hunderfund AN, Martinez-Thompson JM, Kumar N, Savica R, Cutsforth-Gregory JK, et al. A primer on the in-home teleneurologic examination. Neurol Clin Pract 2020 May 21:10.1212/CPJ.0000000000000876. [doi: 10.1212/cpj.0000000000000876]
46. Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. Mayo Clin Proc 2017 Jan;92(1):129-146. [doi: 10.1016/j.mayocp.2016.10.004] [Medline: 27871627]
47. Gardner R, Cooper E, Haskell J, Harris D, Poplau S, Kroth P, et al. Physician stress and burnout: the impact of health information technology. J Am Med Inform Assoc 2019 Feb 01;26(2):106-114. [doi: 10.1093/jamia/ocy145] [Medline: 30517663]
48. Busis NA, Shanafelt TD, Keran CM, Levin KH, Schwarz HB, Molano JR, et al. Burnout, career satisfaction, and well-being among US neurologists in 2016. Neurology 2017 Jan 25;88(8):797-808. [doi: 10.1212/wnl.0000000000003640]
49. Unertl KM, Novak LL, Johnson KB, Lorenzi NM. Traversing the many paths of workflow research: developing a conceptual framework of workflow terminology through a systematic literature review. J Am Med Inform Assoc 2010;17(3):265-273 [FREE Full text] [doi: 10.1136/jamia.2010.004333] [Medline: 20442143]

Abbreviations

APP: advanced practice provider
EMR: electronic medical record
HIPAA: Health Insurance Portability and Accountability Act
MA: medical assistant

Edited by G Eysenbach, R Kukafka; submitted 15.09.20; peer-reviewed by K Nevel, B Clay; comments to author 16.10.20; revised version received 22.10.20; accepted 23.11.20; published 09.12.20.

Please cite as:
Saliba-Gustafsson EA, Miller-Kuhlmann R, Kling SMR, Garvert DW, Brown-Johnson CG, Lestoquoy AS, Verano MR, Yang L, Falco-Walter J, Shaw JG, Asch SM, Gold CA, Winget M
Rapid Implementation of Video Visits in Neurology During COVID-19: Mixed Methods Evaluation
J Med Internet Res 2020;22(12):e24328
URL: http://www.jmir.org/2020/12/e24328/
doi:10.2196/24328
PMID:33245699

Review

COVID-19 Mobile Apps: A Systematic Review of the Literature

Haridimos Kondylakis[1*], MSc, PhD; Dimitrios G Katehakis[2*], MS; Angelina Kouroubali[1*], PhD; Fokion Logothetidis[2*], MBA; Andreas Triantafyllidis[3*], PhD; Ilias Kalamaras[3*], PhD; Konstantinos Votis[3*], PhD; Dimitrios Tzovaras[3*], PhD

[1]Computational Biomedicine Laboratory, Foundation for Research and Technology - Hellas–Institute of Computer Science, Heraklion, Greece
[2]Center for eHealth Applications and Services, Foundation for Research and Technology – Hellas, Institute of Computer Science, Heraklion, Greece
[3]Information Technologies Institute, Centre for Research and Technology - Hellas, Thessaloniki, Greece
[*]all authors contributed equally

Corresponding Author:
Haridimos Kondylakis, MSc, PhD
Computational Biomedicine Laboratory
Foundation for Research and Technology - Hellas–Institute of Computer Science
N. Plastira 100
Heraklion, 70013
Greece
Phone: 30 2810391449
Email: kondylak@ics.forth.gr

Abstract

Background: A vast amount of mobile apps have been developed during the past few months in an attempt to “flatten the curve” of the increasing number of COVID-19 cases.

Objective: This systematic review aims to shed light into studies found in the scientific literature that have used and evaluated mobile apps for the prevention, management, treatment, or follow-up of COVID-19.

Methods: We searched the bibliographic databases Global Literature on Coronavirus Disease, PubMed, and Scopus to identify papers focusing on mobile apps for COVID-19 that show evidence of their real-life use and have been developed involving clinical professionals in their design or validation.

Results: Mobile apps have been implemented for training, information sharing, risk assessment, self-management of symptoms, contact tracing, home monitoring, and decision making, rapidly offering effective and usable tools for managing the COVID-19 pandemic.

Conclusions: Mobile apps are considered to be a valuable tool for citizens, health professionals, and decision makers in facing critical challenges imposed by the pandemic, such as reducing the burden on hospitals, providing access to credible information, tracking the symptoms and mental health of individuals, and discovering new predictors.

(J Med Internet Res 2020;22(12):e23170) doi:10.2196/23170

KEYWORDS

mobile apps; systematic survey; COVID-19; mobile health; eHealth

Introduction

The COVID-19 outbreak, which first emerged in China, has spread worldwide. On March 11, 2020, the World Health Organization (WHO) declared COVID-19 as a pandemic [1]. The disease has disrupted global trade, employment, and travel, and many governments had to take strict measures to control the spread of the virus and minimize the burden of morbidity and mortality so that health care systems remain functional [2]. In many countries around the world, citizens have been recommended to stay at home and practice social distancing for as long as possible as a primary measure of preventing the spread of COVID-19.

Although mobile apps are successfully used for managing chronic diseases [3,4], the ongoing COVID-19 pandemic has pushed the need for mobile app solutions at the forefront to reduce the risk of cross-contamination caused by close contact [5-7]. Mobile technology has been leveraged in a number of ways to control the spread of COVID-19. Mobile apps are accessible, acceptable, and easily adopted, and have the ability to support social distancing efforts. As such, they have been widely developed and implemented during the previous months in an attempt to “flatten the curve” of the increasing number of COVID-19 cases, providing knowledge and information to civilians while attempting to relieve the pressure from health care systems.

Despite increasing reliance on mobile health (mHealth) solutions as part of COVID-19–related response plans, major knowledge gaps exist about their utility and efficacy during the current pandemic for both health professionals as well as for the general population. To this direction, this systematic review aims to shed light into studies found in the scientific literature on the use and evaluation of mobile apps for the prevention, management, treatment, or follow-up of COVID-19.

Other recent reviews have focused merely on the exploration of COVID-19 mobile apps in app stores in general [8] or were restricted to apps deployed in specified countries such as the United States, the United Kingdom, and India [9]. Although there are already related generic COVID-19 information and communication technology surveys [10,11], they focus on specific topics such as contact tracing [12,13]; specialized health sectors like pediatric health care delivery [14], mental health [15], epilepsy [16], and palliative care [17]; or countries like India [14], China [18], and the United Kingdom [19]. To the best of our knowledge, there has been no other work dedicated to the systematic review of pragmatic studies that have demonstrated the real-life use and evaluation of COVID-19 mobile apps.

Methods

Search Strategy

The bibliographic databases of PubMed and Scopus, along with the global research database on COVID-19 developed by the WHO [20], were searched to identify mHealth apps used for the purposes of prevention, treatment, or management of COVID-19 and assessed in pragmatic studies.

Eligibility Criteria

In this context, the inclusion criteria for study selection were the following: features of the COVID-19 mobile app should be described, the study should show evidence of the implementation of the COVID-19 mobile app in real life and provide quantitative outcomes, the study should show that clinical professionals were involved in the design or validation of the mobile app, the paper describing the study must have been written in English. Case reports, letters to editors, preprint papers, qualitative studies, surveys or reviews, simulation studies, and studies describing protocols were excluded from the review.

Study Selection

The string “(mobile health) OR (mhealth) OR (smartphone) OR (mobile phone) OR (mobile application) OR (mobile app) OR (app) AND (COVID-19),” was used for a search within the title, abstract, and keywords of the manuscripts. Authors HK, AT, AK, FL, DGK, and IK independently screened the identified papers to minimize possible errors and bias in the selection process. Any disagreements were resolved by discussion between the authors to reach consensus. The authors first screened the abstracts of the candidate papers for inclusion and assessed their eligibility according to the defined inclusion and exclusion criteria. Moreover, the authors selected the final papers for inclusion after reading the full manuscripts of the eligible papers, as well as their references.

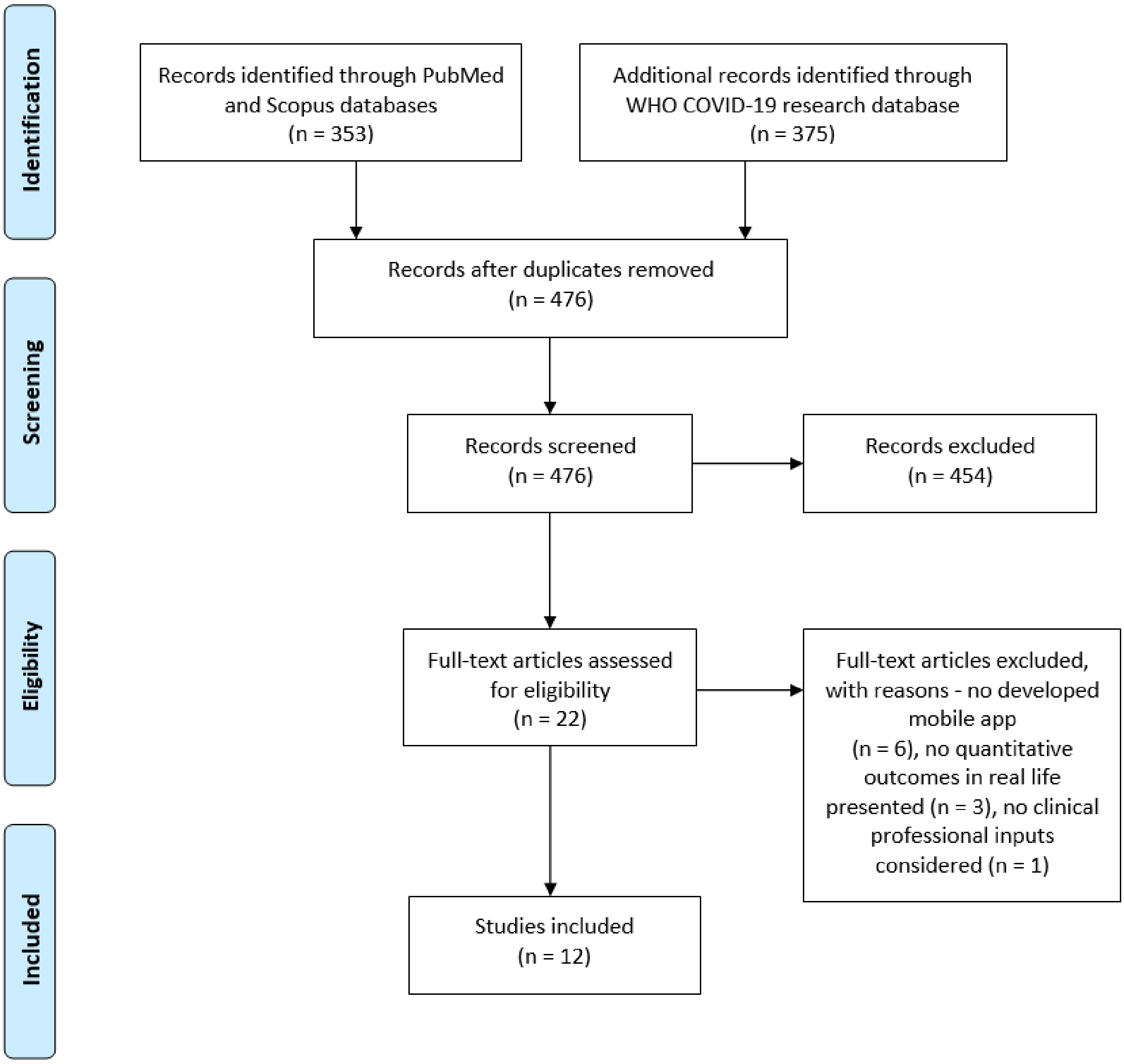
The Effective Public Health Practice Project (EPHPP) tool was adopted to assess the methodological quality of the included studies. The EPHPP tool is suitable for evaluating quantitative studies in a wide range of health-related topics, and it has demonstrated reliability [21]. The included studies were synthesized (by HK, AK, FL, and DGK) according to their target, mobile app main features, study design, number of enrolled participants and their age, follow-up duration, outcomes and whether these were positive or negative, and implications for clinical practice. This systematic review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [22]. A completed PRISMA checklist is shown in Multimedia Appendix 1.

Results

Statistical Analysis

Our last search in June 2020 returned 165 manuscripts from the PubMed database, 188 manuscripts from the Scopus database, and another 375 manuscripts from the WHO research database. All the retrieved records were imported in the Zotero bibliography management software (Center for History and New Media at George Mason University), which identified 252 duplicates. We screened the abstracts of the remaining 476 papers according to our inclusion and exclusion criteria, and 22 papers were found to be eligible. After reading the full text of the papers, the authors agreed to include 12 papers. The screening procedure along with reasons for excluding papers are shown in the PRISMA flow diagram in Figure 1.

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. WHO: World Health Organization.



Study Outcomes and Quality Assessment

On the basis of the EPHPP criteria for selection bias, design, confounders, blinding, data collection, and dropouts, we found the methodological quality to be moderate for 2 of the 12 (17%) studies [23,24] and weak for the remaining 10 (83%) studies (Table 1). Most studies were poorly rated because of their observational or cross-sectional nature, insufficient care in controlling for confounders, insufficient reporting on the validity and reliability of the tools used for data collection, and the absence of description on withdrawals and dropouts. The design of a randomized or controlled clinical trial was not reported in any of the reviewed studies.

Table 1. Quality assessment of included studies based on the EPHPP criteria.

Study	EPHPP[a] criteria						Global rating
	SB[b]	SD[c]	CF[d]	BL[e]	DC[f]	WD[g]	
Bae et al [25]	W[h]	W	W	M[i]	W	W	W
Bourdon et al [23]	M	M	M	M	M	W	M
Drew et al [26]	W	W	W	M	W	W	W
Ben Hassen et al [27]	W	W	W	M	S[j]	W	W
Huckins et al [24]	W	M	M	M	M	S	M
Kodali et al [28]	W	W	W	M	W	W	W
Medina et al [29]	M	W	W	M	W	W	W
Menni et al [30]	W	W	S	M	W	W	W
Ros and Neuwirth [31]	M	W	W	M	W	W	W
Timmers et al [32]	M	M	W	M	W	W	W
Yamamoto et al [33]	M	W	W	M	W	W	W
Zamberg et al [34]	W	W	W	M	W	W	W

[a] EPHPP: Effective Public Health Practice Project.
[b] SB: selection bias.
[c] SD: study design.
[d] CF: confounders.
[e] BL: blinding.
[f] DC: data collection methods.
[g] WD: withdrawals and dropouts.
[h] W: weak.
[i] M: moderate.
[j] S: strong.

Comparison

In the following sections, we provide a comparison of the included studies, in terms of intervention target and main features, study design, and outcome and clinical implications.

Intervention Target and Main Features

The reviewed papers varied in intervention target and main features (Table 2). One paper discussed a COVID-19 tracking app informing users about risk assessment and offering relevant advice [28]. The paper did not provide further details about app features. Apps that offered education and information material were discussed in 3 papers. They either targeted patients or health care professionals. Patient education included hospital information (appointments, visiting hours, self-isolation, and general COVID-19 information [32]) and education related to SARS-CoV-2. For health care professionals, they included information about how to handle patients with COVID-19 [31,34]. One paper [25] described how the hospital information system was adapted to handle COVID-19 with the incorporation of relevant templates. An app for recording general health information was discussed in 1 paper [33], and apps for self-assessment focusing on symptoms were found in 1 paper [32]. Self-monitoring based on a symptom diary and health observation data were also reported in 4 papers [26,29,32,33].

Table 2. A comparison of intervention target and main features.

Studies	Intervention	Main features
Kodali et al [28]	Tracking app	Risk assessment and advice
Timmers et al [32]	Education and information	Information for patients
Ros and Neuwirth [31] and Zamberg et al [34]	Education and information	Information for health care professionals
Bourdon et al [23]	Hospital information system	Adaptation to pandemic
Yamamoto et al [33]	Health assessment	General
Bae et al [25], Medina et al [29], Timmers et al [32], and Yamamoto et al [33]	Health assessment	Self-monitoring
Medina et al [29]	Home monitoring	Telephone outreach
Bourdon et al [23]	Home monitoring	Vital signs with smart devices
Drew et al [26]	Home monitoring	Internet of Things
Timmers et al [32]	Interactive map	Demographics and health status
Yamamoto et al [33]	Data sharing	Patient email to specific recipients
Medina et al [29]	Data sharing	Nurse outreach
Ben Hassen et al [27]	Student mental health	Behavior during the pandemic
Menni et al [30]	Prediction model	Progress of disease
Huckins et al [24]	Teleophtalmology	Emergency eye care

Home monitoring was handled based on telephone outreach [29], smart vital sign monitoring devices [25], and Internet of Things [27]. An interactive map of user demographics and health status was discussed in 1 paper [32]. Two apps offered data sharing capabilities via email to specific recipients based on patient options [33], or as symptoms were worsening, the app sent messages to be assessed by nursing staff to provide appropriate advice [29]. One paper assessed student behavior to determine whether the mental health of participants changed due to the COVID-19 pandemic [24]. One paper used data collected by Drew et al [26] to develop a prediction model about disease progress as well as tracking the disease progress in real time [30]. Finally, 1 paper [23] reported on a smartphone app to manage emergency eye care.

Study Design and Structure of Research

Regarding the *study design*, almost all selected studies were proof of concept, observational studies, or both, aiming to evaluate the practical use, usability, or user satisfaction with the corresponding apps that were developed to support individuals, health care providers, and policy makers during the COVID-19 pandemic (Table 3).

Table 3. A comparison of study design and structure of the research.

Study	Study design	Participants, n	Age	Follow-up duration (study period)
Bourdon et al [23]	Online surveys on patient and medical staff satisfaction with the mobile app and the wearables	12 patients and 24 medical staff	Patient mean age: 25 years; no information for medical staff	No follow-up
Huckins et al [24]	Observational cohort study measuring behaviors through the StudentLife smartphone sensing app	500	Mean: 40.7 (SD 20.3; 0.6-92) years. The number of patients older than 60 years was small.	No follow-up
Bae et al [25]	Observational data collection that helped develop predictive models. Participants that were already enrolled in ongoing epidemiologic studies were approached to use this app.	>2 million users; 75% female	Mean: 41 (range: 18-90) years	The launching of the COVID Symptom Study app occurred in the United Kingdom on March 24, 2020, and in the United States on March 29, 2020. 265,851 individuals were enrolled by March 27, 2020.
Drew et al [26]	Observational study on the COVID Symptom Tracker mobile app	5 hospitalized patients and 5 doctors	Patients range: 45-61 years	No follow-up
Ben Hassen et al [27]	The StudentLife app was used for smartphone mobile sensing. Ecological momentary assessments were used to assess depression and anxiety.	217; 67.8% (n=147) were female	Range: 18-22 years at the time of enrollment	178 (82.0%) students provided data during the Winter 2020 term (January 6 to March 13, 2020).
Kodali et al [28]	Observational study using descriptive statistics and thematic analysis on the mHealth[a] app Arogya Setu.	503 most relevant reviews were identified based on the Google algorithm	Not reported	All reviews that were available publicly and posted in English by the users until April 21, 2020, were included. The start date of app reviews collection was not reported.
Medina et al [29]	Observational cohort study carried out at the Cleveland Clinic, OH, US. It included a self-monitoring app for patient engagement and early intervention.	COVID-19 patients enrolled by May 25, 2020: 1924. Most (85%) patients were enrolled 5 days from symptom onset.	25% (n=483) were older than 60 years, and 3.5% (n=67) were younger than 18 years.	Engagement with MyCare Companion app reached 32%; 25% continued under monitoring for longer than 14 days due to persistent symptoms.
Menni et al [30]	Observational data collection and statistical analysis that helped develop predictive models	Symptoms were reported by 2,450,569 from the UK and 168,293 from the US	Average age for tested positive, tested negative, and not tested: (UK: 41.25, 41.87, and 43.38; US: 41.87, 47.25, and 53.00).	Data analyzed had been collected between March 24 and April 21, 2020.
Ros and Neuwirth [31]	A tutorial feedback survey was conducted. User feedback was requested from health care workers and responders about the presented global public health educational outreach technology.	12,516 users, learners, health care workers, and responders downloaded the app in 1 month.	Not provided	366 replies received during the first 72 h of deploying the survey. During this time period, there were 512 subscribers that had downloaded the app (71.48% response rate).

Study	Study design	Participants, n	Age	Follow-up duration (study period)
Timmers et al [32]	Observational cohort study (based on the data collected at the ETZ[b] hospital), assessed the use of the app as well as its usability. Data were gathered for health care providers and policy makers.	6194 individuals downloaded the app.	Average: 50.87 years	The study focused on data collected between April 1-20, 2020. The app was being used by over 15 hospitals in the Netherlands, Belgium, and Germany, accumulating over 30,000 downloads.
Yamamoto et al [33]	Proof of concept and practical use study in a real-world setting. The study aimed to develop a PHR[c]-based COVID-19 symptom-tracking app to determine whether PHRs could be used for efficient health observation outside a traditional hospital setting. The practical aspects of health observations for COVID-19 using the smartphone or tablet app integrated with PHRs was demonstrated. Moreover, a usability evaluation of the app was carried out based on interviews with help desk managers of the app.	In the context of the active epidemiological investigation period (from March 6-19, 2020) at Wakayama City Public Health Center, 72 individuals who had close contact with a COVID-19 confirmed case were discovered. Among them, 57 had adopted the use of the health observation app.	N/A[d]	The active epidemiological investigation period was carried out from March 6-19, 2020, at Wakayama City Public Health Center. In this period 57 of 72 individuals (health observers) adopted the use of the app. By mid-May, the app had been used by more than 20,280 users and 400 facilities and organizations. These included companies, schools, hospitals, and local governments across Japan.
Zamberg et al [34]	Utilization-focused evaluation study to identify the use of an mHealth platform for information sharing	125 members of the hospital staff	25-30 years: 28 members; 31-35 years: 24 members; 36-40 years: 18 members; 41-50 years: 29 members; 51-60 years: 24 members; >60 years: 2 members	The mHealth platform was used for 18 days from February 25, 2020, until March 13, 2020.

[a] mHealth: mobile health.

[b] ETZ: Elisabeth Twee Steden.

[c] PHR: personal health record.

[d] N/A: not applicable.

One study [24] used the corresponding mobile app (StudentLife app) as a tool for an ecological momentary assessment, aiming to assess depression and anxiety of college students to determine whether there was a change in behavior and mental health of participants in response to COVID-19.

The *number of participants* was mentioned in all selected studies. Some studies reported a relatively low numbers of participants using the corresponding app, whereas others a relatively large number of participants and users. The studies that reported the largest number of users were those described by Drew et al [26] (over 2 million users) and Menni et al [30] (2,450,569 users in the United Kingdom and 168,293 from the United States that had reported symptoms through the smartphone app). Studies that reported the lowest number of participants were the ones described by Bae et al [25] (12 patients and 24 medical staff) and Ben Hassen et al [27] (5 hospitalized patients and 5 doctors).

The *age of participants* was also reported in most of the studies except 3 [28,31,33]. The average age or age range for participants varied depending on the corresponding study.

Concerning the *follow-up duration (study duration)*, 3 out of the 12 selected studies [23,25,27], did not report the study duration. Kodali et al [28] reported the end date but did not mention the start date of app use, Drew et al [26] reported only the start date of the app launch, Ros and Neuwirth [31] reported 3 days of app use, Yamamoto et al [33] reported 14 days, Zamberg et al [34] reported 18 days, Timmers et al [32] reported 20 days, and Huckins et al [24] reported 67 days of app use. Finally, Medina et al [29] mentioned that the engagement with the app reached 32%, whereas 25% of users continued under monitoring for longer than 14 days due to persistent symptoms.

Outcomes

Outcome assessments were varied across studies. An overview is shown in Table 4.

Table 4. A comparison of outcomes for the various studies.

Paper	Primary outcomes	Positive/negative outcomes
Timmers et al [32]	• The information provided by the app satisfied the user needs. • Users indicated the added-value of the symptom tracker diary to be high.	• Successful implementation and use of a COVID-19 app for individuals • An interactive map displayed the data collected through the app. • COVID-19 screening results produced at the hospital were linked to app data. • Health care providers and policy makers could use the data in developing their health care strategy based on the distribution of the reported infection load.
Yamamoto et al [33]	• 72 health observers were identified who were in close contact with a confirmed case. Among them, 57 adopted the app, while 14 used telephone as a means for conducting investigations. • Before the introduction of the app, phone interviews required more than 2 hours and four epidemiological officers for contact tracing. After the introduction of the app, only one epidemiological officer was needed to perform health observations. • The visualization of health observation data improved the investigation efficiency and comprehensiveness.	• The ability of individuals to record health status on a daily basis was an important countermeasure against COVID-19. • The use of the app improved the efficiency and completeness of the investigation process for COVID-19 cases carried out by epidemiological officers.
Zamberg et al [34]	• Three documents related to COVID-19 were made available to medical staff via the mobile platform. • Information was viewed 859 times, which accounted for 35.6% of total document views. • The number of sessions per day increased significantly in the study period (more than doubled) compared with the sessions per day in previous weeks. • Usability evaluation: 70 (83.3%) said it was easy to find information about SARS-CoV-2. On a 10-point Likert scale, the mHealth[a] solution scored 8.5 for time-saving and 7.6 for COVID-19 patient care assurance in daily practice.	• Using the mHealth solution as a communication channel turned out to be effective within the organization for dissemination purposes during the pandemic. • Daily practice was conducted by more confident and better-informed health care professionals.
Kodali et al [28]	• Mixed evidence about the use of the app but mainly optimistic	• Error correction, improved data collection quality, and user privacy should be considered in mHealth apps. • Steps must be taken to ensure the reliability of the information provided by users. Therefore, predicting multiple verification of data entered by users could be critical.
Huckins et al [24]	• With the rise of news relevant to COVID-19, college students spent more time seated, had fewer visits, and showed increases in anxiety and depression. • The authors did not observe a return to baseline over the break, although they observed decreases in stress and depression that paralleled the typical drop after the final examination, suggesting some resilience in the face of COVID-19.	• Mobile apps can be effectively used for tracking the mental health of college students.
Drew et al [26]	• The app captures the dynamics of COVID-19 onset days before traditional measures such as positive tests, hospitalizations, or mortality. • The collection emphasizes the potential usefulness of symptom monitoring in real time to help guide the allocation of resources for testing and treatment, as well as advising for tightening or loosening appropriate measures in specific areas.	• With the participation of groups with underrepresented populations, the study aimed to encourage enrollment of individuals from populations that have traditionally been difficult to recruit. • The study could capture correlations based on individual variations over time, a remarkable advantage over repetitive cross-sectional surveys that introduced significant variation between individuals.

Paper	Primary outcomes	Positive/negative outcomes
Medina et al [29]	- Mobile and home-based interventions were feasible for a wide range of conditions with a related risk of poor outcome from COVID-19. - Approximately 10% of the patients in active monitoring presented symptoms such as shortness of breath that required escalation to a virtual provider. The median time to escalation ranged between 7 and 8 days. - Patients with a pulse oximeter at home escalated a few days earlier due to reduced oxygen saturation measurements before subjective complaints of dyspnea. - 2% of patients in active supervision were eventually admitted, and 3% were readmitted for persistent COVID-19 symptoms or due to complications of other underlying diseases. - 9 patients monitored at home died, either due to complications related to COVID-19 or complications of another underlying disease.	- Mobile engagement platforms have the potential to reduce the need for caregiver communication for patients whose symptoms are mild or persistent, freeing up the health care professionals to focus on patients who need it more.
Menni et al [30]	- Besides more established symptoms such as high fever and a persistent cough, loss of smell and taste were possible prognostic factors for COVID-19. - A combination of symptoms such as anosmia, fatigue, persistent cough, and loss of appetite together could identify individuals with COVID-19.	- Physiological assessments of olfactory and taste function or nucleotide-based testing for SARS-CoV-2 could not be replaced by self-reporting. - The authors did not know if anosmia was acquired before or after other COVID-19 symptoms, or during or after the illness.
Ross and Neuwirth [31]	- This app was considered by the users as appropriate to learn and review skills relevant to COVID-19. - More than 95% of respondents gave a score ≥5 for skills acquisition. - 88% of respondents said it matched their health care needs. - 93% of the respondents stated that the app gave them a better understanding. - 87% of the respondents felt quite or very confident about the execution of the procedures, as shown in the lessons. - 94% of respondents said that this particular COVID-19 training program made them feel ready to care for COVID-19 patients. - 95% of respondents would suggest the application to other users.	Advantage over medical videos: - It allowed the user to live the experience of seeing through a first-person view to learn through the eyes of the expert. - The ability for a health care professional to instantly download locally (in a smart phone) material that can be accessed at any time in real time before, during, or after patient care interactions. - When downloaded, the end user could access and view the tutorial at any time, regardless of network signal issues. - Allowed health care professionals to navigate on their own or to jump to sections that were of greater importance to them
Bae et al [25]	- Mobile app: usefulness showed the highest score, followed by satisfaction and perceived ease of use. - Wearable vital sign monitoring perceived usefulness scored the highest, followed by perceived ease of use and satisfaction. - For carers, there was an overall satisfaction score of 4.10/5.	- During periods of pandemics and disasters, automated exchange of information between health care institutions plays an important role in dealing more efficiently with the problem at hand.
Ben Hassen et al [27]	- Patients and doctors alike accepted the home hospitalization system very well.	- Adjustments should be made for COVID-19 patients safely. - Vital signs had to be measured by the patients themselves. - Video communication between patients and doctors was added.
Bourdon et al [23]	- Allowed doctors and patients to maintain social distance, avoiding three or four physical trips per person.	- A physical appointment followed 27% of the teleconsultations. - Average delay of 4.12 days between the onset of symptoms and advice, and <1 day for emergency episodes. - There was 96% sensitivity and 95% specificity for the correct evaluation of the indication of a physical consultation and only 1.0% misdiagnoses.

[a] mHealth: mobile health.

In most of the cases, app users were satisfied with the educational and risk assessment information provided in the mobile apps (both citizens [32] and health personnel [31,34]). They felt reassured and informed. However, there were also concerns [28] about the measures that would need to be taken to ensure the reliability of information provided to users.

In addition, 3 studies [26,30,32] showed the feasibility of the implementation and use of COVID-19 apps to support education, self-management, and symptom tracking through diaries. Symptom tracker apps had even identified potential predictors of COVID-19, such as loss of smell and taste [30]. To this direction, mobile app sensing revealed that, during the COVID-19 pandemic, individuals were more sedentary, visited fewer locations, and showed increases in anxiety and depression [24], which is in line with the "Stay safe, stay home" policies mandated by local and national governments. However, there were concerns that cross-verification of data entered by the users could be crucial [28] and that the self-reported nature of the data collected by these apps could not replace physiological assessments and clinical examinations [30].

Additionally, mobiles apps were used for home monitoring, as reported in 4 studies [23,25,27,29], with satisfactory usability for both patients and health care professionals. These apps were focused on the management of patients who are at high risk, the moderation of exposure risk for health care workers, and the reductions of community spread through appropriate education on home-based care for individuals who are exposed or infected.

Based on the data reported in the reviewed papers, interactive maps and dashboards could be created for the quick visualization of the status summary of patients with COVID-19 and disease spread to be used by policy makers and health care providers for decision making at regional levels [26,32,33]. In addition, the dynamics of incidence days could be effectively captured, guiding allocation of resources for testing, treatment, and lockdown recommendations [26]. Mobile apps could significantly improve the efficiency and completeness of contact tracing workflows [33].

Discussion

Main Findings

A systematic review on COVID-19 mobile apps, as used and evaluated in research studies published in the scientific literature, is presented. Our literature search returned a significant number of records (476 unique published manuscripts), despite the short time period covered (December 2019 to June 2020), thereby showing the high interest of the scientific community in the research of mobile apps for COVID-19.

Our main finding is that, despite that the current research evidence is fragmented and requires greater methodological rigor, mobile apps have been found to benefit citizens, health professionals, and decision makers in facing the COVID-19 pandemic. In particular, mobile apps can help in solving several COVID-19–related challenges by increasing the reach of reliable information to both citizens and health professionals, decreasing misinformation and confusion, tracking symptoms and mental health of citizens, home monitoring and isolation, discovering new predictors, optimizing health care resource allocation, and reducing the burden of hospitals.

The participants in the studies were mainly young and middle-aged adults. Further studies are needed that will involve older participants, who are in greater risk of developing serious complications due to COVID-19. Understanding the needs of older individuals in the COVID-19 pandemic period would be the first necessary step toward designing and developing mobile apps to encourage their physical and mental well-being [17].

Our review, in contrast to other reviews that have not examined the evaluation of COVID-19 mobile apps in pragmatic studies [8], identified that the majority of included studies were not of high methodological quality, mainly because of their observational nature. This could be justified by the fact that the COVID-19 pandemic crisis generated an international appeal for fast response and rapid development of digital health tools by the research community, which might have inevitably led to the publication of early results by observational studies. This result can be seen as complementary to other reviews [8] that report that many of those apps are of high quality, offering many functionalities and advanced user experience. Longitudinal studies with rigorous design such as randomized controlled trials are now required to systematically assess COVID-19 mobile apps and provide strong evidence of their value. However, ethical implications might arise due to possible conflicts between liberty and privacy, equity, fairness, and justice [35-37]. In this direction, health outcomes that have scarcely been used so far, such as infection rate and quality of life, could be used as primary end points.

Implications for Clinical Practice

The implications for clinical practice for each one of the discussed works is shown in detail in Textbox 1.

Textbox 1. Implications for clinical practice.

Timmers et al [32]

- eHealth mobile apps for COVID-19 that support several functionalities such as tracking of symptoms and provision of accurate and timely information as well as self-assessment can be implemented in a short time to be used by individuals.
- Those apps provide valuable information to both governments and health care providers, since they support monitoring of patient health status and provide summary statistics regarding the progress of health.
- Such apps could be used in future outbreaks of other viruses to support all involved stakeholders.

Yamamoto et al [33]

- Such apps could be used in future outbreaks of other viruses to support all involved stakeholders.

Zamberg et al [34]

- Mobile health (mHealth) apps could help solve some of the COVID-19 challenges by providing more accurate and timely information to health care professionals. This benefit is sourced by the centralized management and storage of main, up-to-date, validated, and easily accessed information in one platform.
- mHealth apps used in health care organizations as communication tools of validated information should be assessed in the context of clinical studies, with regards to their impact on clinical outcomes.

Kodali et al [28]

- According to users, the app should be enhanced with additional functionalities including tracking of location, provision of up-to-date information on COVID-19 as well as information on areas with high/medium/low epidemiological burden, and deployment in nonmobile platforms.

Huckins et al [24]

- According to the findings, since the beginning of the COVID-19 pandemic, there has been an extended negative impact both on the physical health of individuals, resulting in an increased number of deaths, and in their mental health, resulting in changes in their behavior.

Drew et al [26]

- The research of a broader range of potential risk factors for COVID-19 results will be largely supported by the use of the app within several large epidemiology cohorts for which there are a large amount of data on lifestyle, diet and health factors, and genetic information.
- It will be useful to deploy the tool in several clinical studies, in centralized actions related to biobanking, and in health care worker monitoring programs.

Medina et al [29]

- The intervention impacts several organizational matters and provides answers regarding the following questions related to COVID-19:
    - Which situations should we focus on to provide effective care to prevent patient admission to intensive care unit and mechanical support?
    - How can those situations be better supported by increased patient self-reporting?
    - Can we predict, based on time data, future inpatient demand?
    - How can we manage not having conflicts between patient choice and availability of treatment?

Menni et al [30]

- Routine screening for COVID-19 could also include loss of sense of taste and smell. These two symptoms should also be included in the related symptom list provided by the World Health Organization.

Ros and Neuwirth [31]

- The implementation and deployment of the described digital tutorials, as an effective and swift global public health educational tool, help alleviate the burden that hospitals, health care professionals/responders, and patients face due to the COVID-19 pandemic.

Bae et al [25]

- Episode triage, timely diagnosis, isolation of patients, and their treatment can be largely supported by telemedicine solutions.

Ben Hassen et al [27]

- Mobile apps can be used by patients, their caregivers, and health care professionals to better monitor and manage patient health status in the context of patient hospitalization. Moreover, such apps are cost-effective, reliable, and safe, providing important economic benefits to hospitals. Overall, they are accepted to a great extent by patients and individuals.

Bourdon et al [23]

- The use of telemedicine tools for consultation purposes improved the access to health care services for patients with ophthalmological problems. At the same time, such tools preserved social distancing and sanitary measures.
- Such tools can be used in emergency situations by ophthalmological patients who have limited access to specialized care.

As reported in the textbox, eHealth solutions can be implemented rapidly and can offer essential tools in supporting the COVID-19 pandemic for all stakeholders including citizens, health care providers, policy makers, and governments. The impact of the COVID-19 pandemic goes beyond the illness and deaths that are directly associated with the SARS-CoV-2 pathogen and toward an expanded scope encompassing mental health and several behavioral changes.

However, there are still several areas to explore, such as better geolocation tracking; timely COVID-19 updates; deployment on nonsmartphone platforms; better incidence visualization and prediction; investigation of a much broader range of putative risk factors for COVID-19 outcomes (based on other longitudinal data on lifestyle, diet, genetics, etc); and the integration with clinical systems, digital health literacy, and engagement with mobile apps.

Apps such as the ones presented in these studies can give the research community the opportunity to monitor long term effects of COVID-19; answer questions such as the number of people truly affected, why some people get sicker than others, and how long people can stay immune to the disease; and perform better resource allocation of medical equipment.

In case of future disease outbreaks, the mobile apps already developed for COVID-19 can be valuable tools ready to support people, health care providers, and policy makers. This study can provide a guide for future developers and researchers regarding the current methodological gaps and challenges that need to be addressed to develop well-designed and evaluated apps for similar future circumstances.

Limitations

We used terms related to mobile apps, mHealth, and mobile phones for our literature search according to the review objectives and did not use related terms such as “telehealth,” which might have resulted in an inadvertent omission of studies. Our literature search was conducted in a limited number of bibliographic databases (Pubmed, Scopus, and the WHO’s COVID-19 database), which nevertheless have been largely used worldwide. The gray literature was not explored. The interrater reliability between the authors was not assessed. A meta-analysis was not possible due to the heterogeneity of the included studies. For many cases, the quality of the study might not reflect the quality or the effectiveness of the developed mobile app. As such, useful and effective apps might have not been included in this review study due to the limited quality of the related studies.

Acknowledgments

This research was supported by the Computational Biomedicine Lab and the Center for eHealth Applications and Services of the Foundation for Research and Technology – Hellas, and the Information Technologies Institute of the Centre for Research and Technology – Hellas.

Work presented in the paper is part of the KRIPIS project “Novel technologies to monitor, support and improve the quality of life of patients and other vulnerable user groups (Quality of Life)” funded in the context of the National Strategic Reference Framework 2014-2020, European Regional Development Fund.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 checklist.
[DOC File , 66 KB - jmir_v22i12e23170_app1.doc]

References

1. Coronavirus disease 2019 (COVID-19): situation report – 78. World Health Organization. 2020 Apr 07. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200407-sitrep-78-covid-19.pdf [accessed 2020-07-01]
2. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global preparedness against COVID-19: we must leverage the power of digital health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
3. Triantafyllidis A, Kondylakis H, Votis K, Tzovaras D, Maglaveras N, Rahimi K. Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: A review of systematic reviews. Int J Med Inform 2019 Dec;132:103984. [doi: 10.1016/j.ijmedinf.2019.103984] [Medline: 31605884]

4. Kouroubali A, Koumakis L, Kondylakis H, Katehakis DG. An integrated approach towards developing quality mobile health apps for cancer. In: Moumtzoglou A, editor. Mobile Health Applications for Quality Healthcare Delivery. Hershey, PA: IGI Global; 2019:46-71.
5. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Jun;26(5):309-313 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]
6. Kouroubali A, Kondylakis H, Kavlentakis G, Logothetides F, Stathiakis N, Petrakis Y, et al. An eHealth platform for the holistic management of COVID-19. Stud Health Technol Inform 2020 Sep 04;273:182-188. [doi: 10.3233/SHTI200636] [Medline: 33087610]
7. Katehakis D, Kavlentakis G, Stathiakis N, Logothetidis F, Kouroubali A, Kondylakis H, et al. An outbreak response tool to effectively support surveillance of suspect, probable and confirmed incidence cases while staying safe in COVID-19. 2020 Presented at: 20th IEEE International Conference on Bioinformatics and Bioengineering; October 2020; Virtual conference.
8. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]
9. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
10. Chamola V, Hassija V, Gupta V, Guizani M. A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. IEEE Access 2020;8:90225-90265. [doi: 10.1109/access.2020.2992341]
11. Iyengar K, Upadhyaya G, Vaishya R, Jain V. COVID-19 and applications of smartphone technology in the current pandemic. Diabetes Metab Syndr 2020;14(5):733-737 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.033] [Medline: 32497963]
12. García-Iglesias JJ, Martín-Pereira J, Fagundo-Rivera J, Gómez-Salgado J. [Digital surveillance tools for contact tracking of infected persons by SARS-CoV-2]. Rev Esp Salud Publica 2020 Jun 23;94:e202006067 [FREE Full text] [Medline: 32572019]
13. Ekong I, Chukwu E, Chukwu M. COVID-19 mobile positioning data contact tracing and patient privacy regulations: exploratory search of global response strategies and the use of digital tools in Nigeria. JMIR mHealth uHealth 2020 Apr 27;8(4):e19139 [FREE Full text] [doi: 10.2196/19139] [Medline: 32310817]
14. Badawy SM, Radovic A. Digital approaches to remote pediatric health care delivery during the COVID-19 pandemic: existing evidence and a call for further research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]
15. Merchant R, Torous J, Rodriguez-Villa E, Naslund J. Digital technology for management of severe mental disorders in low-income and middle-income countries. Curr Opin Psychiatry 2020 Sep;33(5):501-507. [doi: 10.1097/YCO.0000000000000626] [Medline: 32520747]
16. Santos-Peyret A, Durón RM, Sebastián-Díaz MA, Crail-Meléndez D, Goméz-Ventura S, Briceño-González E, et al. [E-health tools to overcome the gap in epilepsy care before, during and after COVID-19 pandemics]. Rev Neurol 2020 May 01;70(9):323-328 [FREE Full text] [doi: 10.33588/rn.7009.2020173] [Medline: 32329045]
17. Kunz R, Minder M. COVID-19 pandemic: palliative care for elderly and frail patients at home and in residential and nursing homes. Swiss Med Wkly 2020 Mar 23;150:w20235. [doi: 10.4414/smw.2020.20235] [Medline: 32208497]
18. Pan X. Application of personal-oriented digital technology in preventing transmission of COVID-19, China. Ir J Med Sci 2020 Nov;189(4):1145-1146 [FREE Full text] [doi: 10.1007/s11845-020-02215-5] [Medline: 32219674]
19. Chidambaram S, Erridge S, Kinross J, Purkayastha S, PanSurg Collaborative. Observational study of UK mobile health apps for COVID-19. Lancet Digit Health 2020 Aug;2(8):e388-e390 [FREE Full text] [doi: 10.1016/S2589-7500(20)30144-8] [Medline: 32835196]
20. COVID-19: global literature on coronavirus disease. World Health Organization. 2020. URL: https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/ [accessed 2020-11-01]
21. Armijo-Olivo S, Stiles C, Hagen N, Biondo P, Cummings G. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
23. Bourdon H, Jaillant R, Ballino A, El Kaim P, Debillon L, Bodin S, et al. Teleconsultation in primary ophthalmic emergencies during the COVID-19 lockdown in Paris: experience with 500 patients in March and April 2020. J Fr Ophtalmol 2020 Sep;43(7):577-585 [FREE Full text] [doi: 10.1016/j.jfo.2020.05.005] [Medline: 32564983]

24. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
25. Bae YS, Kim KH, Choi SW, Ko T, Jeong CW, Cho B, et al. Information technology-based management of clinically healthy COVID-19 patients: lessons from a living and treatment support center operated by Seoul National University Hospital. J Med Internet Res 2020 Jun 12;22(6):e19938 [FREE Full text] [doi: 10.2196/19938] [Medline: 32490843]
26. Drew D, Nguyen L, Steves C, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]
27. Ben Hassen H, Ayari N, Hamdi B. A home hospitalization system based on the Internet of Things, Fog computing and cloud computing. Inform Med Unlocked 2020;20:100368 [FREE Full text] [doi: 10.1016/j.imu.2020.100368] [Medline: 32537483]
28. Kodali P, Hense S, Kopparty S, Kalapala G, Haloi B. How Indians responded to the Arogya Setu app? Indian J Public Health 2020 Jun;64(Supplement):S228-S230 [FREE Full text] [doi: 10.4103/ijph.IJPH_499_20] [Medline: 32496261]
29. Medina M, Babiuch C, Card M, Gavrilescu R, Zafirau W, Boose E, et al. Home monitoring for COVID-19. Cleve Clin J Med 2020 Jun 11:1 [FREE Full text] [doi: 10.3949/ccjm.87a.ccc028] [Medline: 32409432]
30. Menni C, Valdes A, Freidin M, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040. [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
31. Ros M, Neuwirth L. Increasing global awareness of timely COVID-19 healthcare guidelines through FPV training tutorials: portable public health crises teaching method. Nurse Educ Today 2020 May 21;91:104479 [FREE Full text] [doi: 10.1016/j.nedt.2020.104479] [Medline: 32473497]
32. Timmers T, Janssen L, Stohr J, Murk JL, Berrevoets MAH. Using eHealth to support COVID-19 education, self-assessment, and symptom monitoring in the Netherlands: observational study. JMIR mHealth uHealth 2020 Jun 23;8(6):e19822 [FREE Full text] [doi: 10.2196/19822] [Medline: 32516750]
33. Yamamoto K, Takahashi T, Urasaki M, Nagayasu Y, Shimamoto T, Tateyama Y, et al. Health observation app for COVID-19 symptom tracking integrated with personal health records: proof of concept and practical use study. JMIR mHealth uHealth 2020 Jul 06;8(7):e19902 [FREE Full text] [doi: 10.2196/19902] [Medline: 32568728]
34. Zamberg I, Manzano S, Posfay-Barbe K, Windisch O, Agoritsas T, Schiffer E. A mobile health platform to disseminate validated institutional measurements during the COVID-19 outbreak: utilization-focused evaluation study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18668 [FREE Full text] [doi: 10.2196/18668] [Medline: 32250958]
35. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
36. Crico C, Renzi C, Graf N, Buyx A, Kondylakis H, Koumakis L, et al. mHealth and telemedicine apps: in search of a common regulation. Ecancermedicalscience 2018;12:853 [FREE Full text] [doi: 10.3332/ecancer.2018.853] [Medline: 30079115]
37. Kondylakis H, Koumakis L, Hänold S, Nwankwo I, Forgó N, Marias K, et al. Donor's support tool: enabling informed secondary use of patient's biomaterial and personal data. Int J Med Inform 2017 Jan;97:282-292. [doi: 10.1016/j.ijmedinf.2016.10.019] [Medline: 27919386]

Abbreviations

EPHPP: Effective Public Health Practice Project
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization

Edited by G Fagherazzi; submitted 03.08.20; peer-reviewed by I Zamberg, P Grace-Farfaglia; comments to author 09.09.20; revised version received 18.09.20; accepted 11.10.20; published 09.12.20.

Please cite as:
Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, Votis K, Tzovaras D
COVID-19 Mobile Apps: A Systematic Review of the Literature
J Med Internet Res 2020;22(12):e23170
URL: http://www.jmir.org/2020/12/e23170/
doi:10.2196/23170
PMID:33197234

Original Paper

Use and Cost-Effectiveness of a Telehealth Service at a Centralized COVID-19 Quarantine Center in Taiwan: Cohort Study

Yung-Feng Yen[1,2,3,4*], MD, PhD; Yi-Fan Tsai[5], MS; Vincent Yi-Fong Su[6,7,8], MD, PhD; Shang-Yih Chan[2,4,6*], MD; Wen-Ruey Yu[8], MD; Hsuan Ho[4], MS; Chun-Mei Hou[4], MS; Chu-Chieh Chen[2], PhD; Lin-Chung Woung[2,9], MD; Sheng-Jean Huang[10,11], MD

[1]Section of Infectious Diseases, Taipei City Hospital, Yangming Branch, Taipei City, Taiwan
[2]Department of Health Care Management, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan
[3]Institute of Public Health, National Yang-Ming University, Taipei, Taiwan
[4]University of Taipei, Taipei, Taiwan
[5]Department of Nursing, Taipei City Hospital Yangming Branch, Taipei, Taiwan
[6]Department of Internal Medicine, Taipei City Hospital Yangming Branch, Taipei, Taiwan
[7]School of Medicine, National Yang-Ming University, Taipei, Taiwan
[8]Department of Chest Medicine, Taipei Veterans General Hospital, Taipei, Taiwan
[9]Department of Ophthalmology, Taipei City Hospital, Taipei, Taiwan
[10]Department of Neurosurgery, Taipei City Hospital, Taipei, Taiwan
[11]Department of Surgery, Medical College, National Taiwan University Hospital, Taipei, Taiwan
*these authors contributed equally

Corresponding Author:
Shang-Yih Chan, MD
Department of Internal Medicine
Taipei City Hospital Yangming Branch
No.145 Zhengzhou Road
Datong District
Taipei, 103
Taiwan
Phone: 886 2 2555 3000
Email: dad89@tpech.gov.tw

Abstract

Background: Telehealth is a recommended method for monitoring the progression of nonsevere infections in patients with COVID-19. However, telehealth has not been widely implemented to monitor SARS-CoV-2 infection in quarantined individuals. Moreover, studies on the cost-effectiveness of quarantine measures during the COVID-19 pandemic are scarce.

Objective: In this cohort study, we aimed to use telehealth to monitor COVID-19 infections in 217 quarantined Taiwanese travelers and to analyze the cost-effectiveness of the quarantine program.

Methods: Travelers were quarantined for 14 days at the Taiwan Yangmingshan quarantine center and monitored until they were discharged. The travelers' clinical symptoms were evaluated twice daily. A multidisciplinary medical team used the telehealth system to provide timely assistance for ill travelers. The cost of the mandatory quarantine was calculated according to data from the Ministry of Health and Welfare of Taiwan.

Results: All 217 quarantined travelers tested negative for SARS-CoV-2 upon admission to the quarantine center. During the quarantine, 28/217 travelers (12.9%) became ill and were evaluated via telehealth. Three travelers with fever were hospitalized after telehealth assessment, and subsequent tests for COVID-19 were negative for all three patients. The total cost incurred during the quarantine was US $193,938, which equated to US $894 per individual.

Conclusions: Telehealth is an effective instrument for monitoring COVID-19 infection in quarantined travelers and could help provide timely disease management for people who are ill. It is imperative to screen and quarantine international travelers for SARS-CoV-2 infection to reduce the nationwide spread of COVID-19.

(J Med Internet Res 2020;22(12):e22703) doi:10.2196/22703

KEYWORDS

COVID-19; international travelers; quarantine; telehealth; cost-effectiveness; cohort; monitoring; telemedicine

Introduction

COVID-19 is caused by SARS-CoV-2 and was first detected in Wuhan, China, in December 2019. This disease caused a rapidly accelerating global pandemic in 2020 [1]; as of November 13, 2020, more than 51.8 million individuals were infected with COVID-19 globally, with the official death toll reaching 1.3 million [2].

As the COVID-19 outbreak emerged, the Taiwanese government implemented several strategies to prevent the nationwide spread of COVID-19, including border controls, proactive screening measures, and quarantine procedures [3,4]. According to the Taiwan Communicable Disease Control Act [5], as of January 2020, all international travelers who visited regions with a declared COVID-19 outbreak were required to complete a 14-day mandatory quarantine. By November 13, 2020, 597 laboratory-confirmed COVID-19 cases were reported to the Taiwan Centers for Disease Control, including 505 imported cases (84.6%) [6]. The national mortality rate among laboratory-confirmed COVID-19 cases was 1.2% [6].

Currently, no effective pharmacological interventions or vaccines are available to treat or prevent COVID-19 [7,8]. Therefore, nonpharmacological public health measures such as quarantine, isolation, social distancing, and community containment are the only effective ways to prevent infection and control the COVID-19 outbreak [9]. Quarantine is the most effective tool for controlling the COVID-19 outbreak [9]; it refers to the restriction of asymptomatic healthy people who have had contact with confirmed or suspected COVID-19 cases. Quarantine can be voluntary or mandatory, and it can be applied at an individual or group level. A recent Cochrane review [10] showed that the quarantine strategy could significantly reduce the number of people infected with SARS-CoV-2 and decrease the number of COVID-19–related deaths.

During quarantine, all individuals should be monitored for the onset of symptoms; otherwise, there is risk of delays in the detection and prompt management of the virus. Telehealth through a line of communication is an efficient way to monitor quarantined individuals and can assist in providing timely care for people who require it [11]. Although telehealth has been recommended to screen patients with COVID-19 in emergency departments [12] and provide care for nonsevere COVID-19 cases [13], it has not been widely implemented in the monitoring of COVID-19 infection in quarantined individuals.

Quarantine strategies have been used in many countries in an attempt to curb the ongoing COVID-19 pandemic. However, the cost-effectiveness of these strategies has not been extensively studied [10]. A current report [14] indicates that the quarantine strategy is efficient in curbing the spread of COVID-19, and the cost of a quarantine strategy is lower than that of lockdown of workplaces.

This cohort study reports on a model used for screening and quarantining Taiwanese travelers following an evacuation flight from Hubei, China. In this model, telehealth was used to monitor COVID-19 infection among travelers at a centralized quarantine center, and the cost-effectiveness of the 14-day mandatory quarantine strategy was analyzed.

Methods

Background Information and Study Subjects

In late January 2020, the Taiwanese government established a centralized quarantine center to screen and monitor COVID-19 infections in international travelers who had visited countries with a declared COVID-19 outbreak [5]. On March 30, 2020, the government of Taiwan arranged a special charter flight to evacuate Taiwanese travelers who had become stranded in Hubei, China, and return them to Taiwan. Upon arrival in Taiwan, the travelers proceeded to one of the largest national centralized quarantine centers, Taipei Yangmingshan, for the mandatory 14-day quarantine period. The Yangmingshan quarantine center included six quarantined wards on three floors.

This cohort study included all the passengers of the charter flight, who were subsequently followed up until their discharge from the quarantine center or until April 15. This study was approved by the Institutional Review Board of Taipei City Hospital (no. TCHIRB-10904014-E).

Infection Control Strategies

The body temperature of the Taiwanese travelers was checked at the international airport in China, and individuals with fever were not allowed to board the charter flight. All travelers were required to disinfect their hands with alcohol-based hand sanitizer at the boarding port and to wear personal protective equipment (PPE) provided by the Taiwanese government, which included a face mask and medical gown. Meals were not provided during the flight to avoid cross-infection between passengers.

Upon arrival at the international airport in Taiwan, National Defense chemical troops used a 1:10 diluted sodium hypochlorite solution (5000 ppm) to disinfect the travelers' luggage. The travelers then proceeded to the Yangmingshan quarantine center for the 14-day quarantine period. They were required to record their body temperature twice daily with a thermometer that was provided to them. Additionally, daily meals were delivered to the individuals' rooms, and janitors used a 1:10 diluted sodium hypochlorite solution to clean the quarantine center once daily.

COVID-19 Screening and Monitoring

During admission to the Yangmingshan centralized quarantine center, the travelers were screened for COVID-19 using a real-time reverse transcriptase–polymerase chain reaction (RT-PCR) test. If an individual tested positive for COVID-19, they were transferred to a hospital for medical isolation and treatment.

Travelers' clinical symptoms were recorded twice daily by their primary care nurses via telephonic communication. If an individual developed fever or became ill, a multidisciplinary team (MDT) consisting of 15 medical workers used telehealth to assess the traveler's condition.

Telehealth System

The telehealth system at the Yangmingshan quarantine center was developed based on a popular social media app called LINE, which is a freeware app operated by the NAVER Corporation in South Korea. The LINE-based telehealth system established two-way communication between the MDT and 217 quarantined travelers. Through a unique ID, the quarantined individuals were invited to join an official LINE group in which the quarantined travelers could report their symptoms to the MDT members. When the MDT members received a message regarding a quarantined traveler's symptoms, the MDT members initiated a one-to-one video call through LINE to clinically evaluate the patient. The MDT members collaboratively discussed the traveler's condition during the telehealth session and provided appropriate management and treatment interventions for the traveler. If a quarantined traveler presented with fever or shortness of breath, they were transferred to the hospital for further treatment.

Personnel and Cost at Yangmingshan Quarantine Center

The personnel who were responsible for overseeing the travelers' quarantine at the Yangmingshan quarantine center included the MDT, police officers, janitors, a logistics group, and administration staff. The MDT consisted of 3 physicians, 10 nurses, a pharmacist, and a psychologist. Three physicians, including a general practitioner, a pediatrician, and a cardiologist, gave treatment advice according to the traveler's condition during the telehealth session. The 10 nurses were responsible for taking care of the travelers during 12-hour day and night shifts in the six quarantined wards. The nurses communicated daily with the quarantined travelers to evaluate their clinical condition and record their body temperature. The pharmacist delivered medications to ill quarantined travelers. Physicians and nurses wore PPE when examining the travelers in person.

The police officers were responsible for the security of the travelers at five sentry points and ensured that all travelers obeyed the COVID-19 quarantine rules and completed the 14-day mandatory quarantine order. In total, 18 police officers were divided into three groups, with one leader in each group. The three groups of police officers worked in 12-hour day and night shifts. Three janitors were responsible for environmental disinfection in the six quarantined wards, and one janitor was responsible for two quarantined wards on one floor. The logistics group, composed of 10 people, was responsible for meal preparation and ensuring that sufficient supplies were provided to the 217 travelers during the 14-day quarantine.

The cost of the quarantine for the travelers at the Yangmingshan quarantine center was covered by the disaster reserve of the national Ministry of Health and Welfare, which included payment of the MDT, nonmedical personnel, and telehealth services as well as the provision of PPE.

Statistical Analysis

First, the participants' demographic data were analyzed. Continuous data are presented as mean (SD), and two-sample *t* tests were used for comparisons between groups. Categorical data were analyzed with Pearson chi-square tests where appropriate. A P value <.05 was considered to indicate statistical significance.

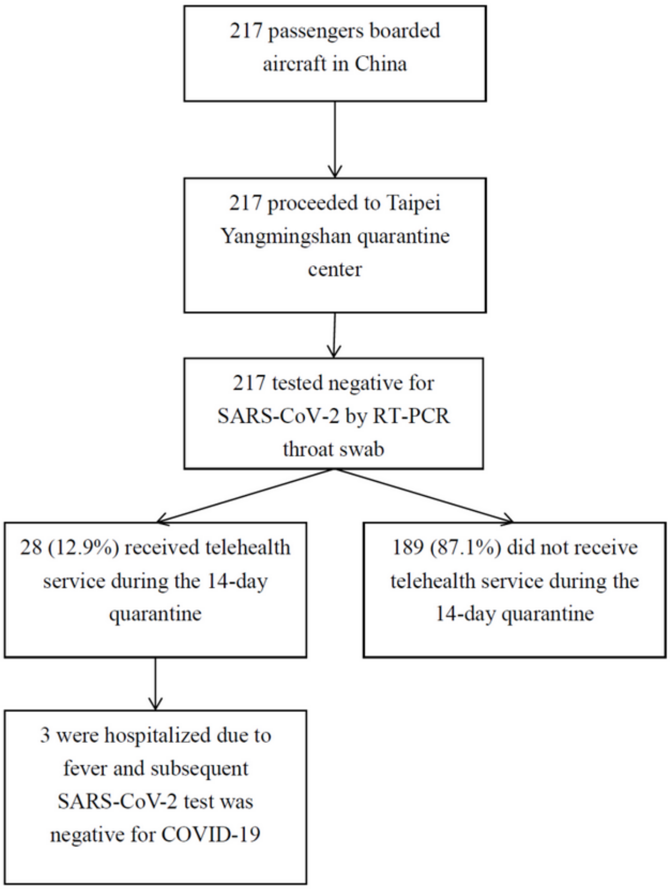
A timeline infographic was used to display the progression of clinical symptoms in hospitalized travelers. The labor and overall costs of the quarantine were calculated according to the disaster reserve data from the Taiwan Ministry of Health and Welfare. All data management and analyses were performed using SPSS 19.0 (IBM Corporation).

Results

Participant Selection and Epidemiologic Features

Our sample comprised 217 Taiwanese travelers who were stranded in Hubei, China, due to the COVID-19 pandemic and who returned to Taiwan on March 30, 2020 (Figure 1). The mean age of the sample was 30.0 years (SD 19.4), and 130 of the 217 subjects (59.9%) were female. All travelers tested negative for COVID-19 by an RT-PCR throat swab upon admission to the Taipei Yangmingshan quarantine center. During the 14-day quarantine, 28 of the 217 travelers (12.9%) underwent telehealth consultations due to illness.

Figure 1. The process of follow-up in Taiwanese travelers quarantined at the Yangmingshan centralized quarantine center. RT-PCR: reverse transcriptase–polymerase chain reaction.



Characteristics of Quarantined Travelers With and Without Telehealth Evaluation

Table 1 shows the characteristics of quarantined travelers who were and were not evaluated with telehealth. There was no significant difference in the age or sex of the quarantined travelers who were and were not evaluated with telehealth. Three travelers had fever and underwent clinical assessment using telehealth during the 14-day quarantine period.

Table 1. Characteristics of Taiwanese travelers who were and were not evaluated with telehealth at a centralized quarantine center.

Characteristic	Total (N=217)	Evaluated with telehealth (n=28)	Not evaluated with telehealth (n=189)	*P* value
Age (years), mean (SD)	30.0 (19.40)	33.0 (18.1)	29.6 (17.3)	0.33
Age (years), n (%)				0.68
<20	78 (35.9)	8 (28.6)	70 (37.0)	
20-39	54 (24.9)	8 (28.6)	46 (24.3)	
≥40	85 (39.2)	12 (42.8)	73 (38.7)	
Sex, n (%)				0.61
Female	130 (59.9)	10 (35.7)	77 (40.7)	
Male	87 (40.1)	18 (64.3)	112 (59.3)	
Number of family members, n (%)				<.001
1	79 (36.4)	19 (67.9)	60 (31.7)	
≥2	138 (63.6)	9 (32.1)	129 (68.3)	
Fever during the 14-day quarantine, n (%)				<.001
No	214 (98.6)	25 (89.3)	189 (100)	
Yes	3 (1.4)	3 (10.7)	0 (0)	
Hospitalization, n (%)				<.001
No	214 (98.6)	25 (89.3)	189 (100)	
Yes	3 (1.4)	3 (10.7)	0 (0)	

Symptoms and Management in Quarantined Travelers Receiving Telehealth Care

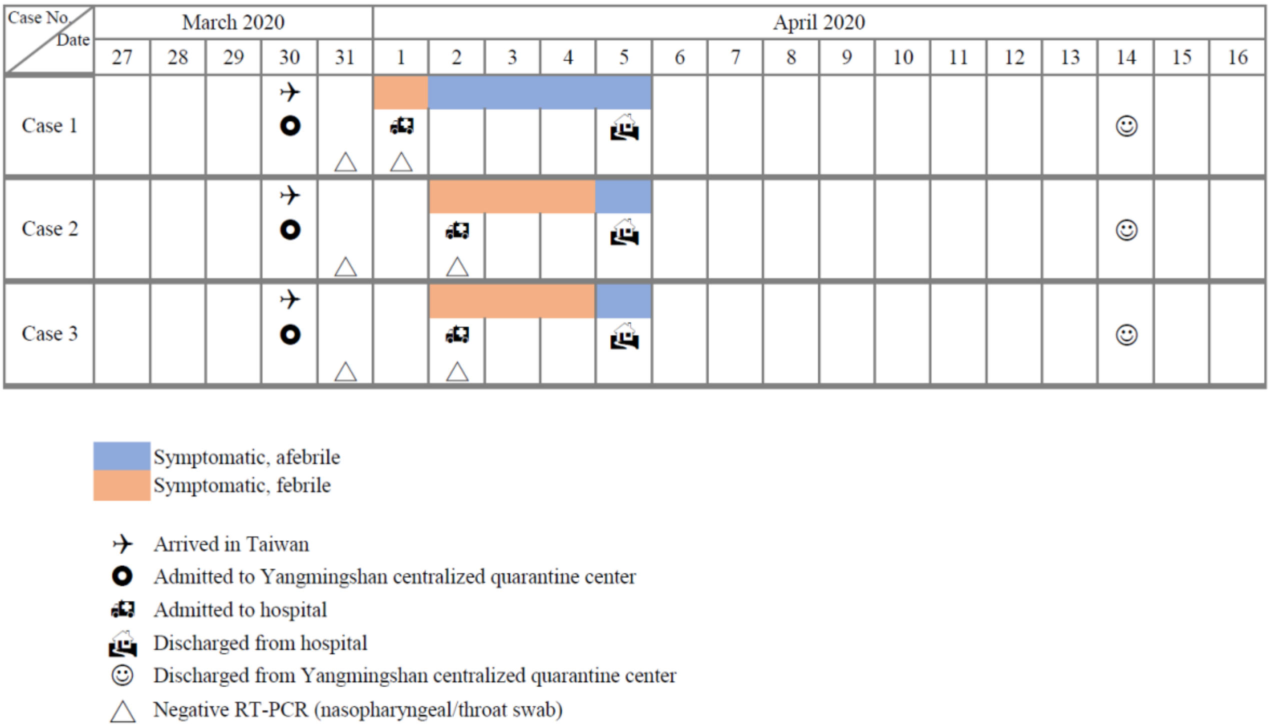
The symptoms and management of quarantined travelers receiving telehealth care were recorded during the 14-day mandatory quarantine (Multimedia Appendix 1). The most common symptoms requiring telehealth consultations in quarantined travelers were fever (n=3), diarrhea (n=3), toothache (n=3), and skin rashes (n=3). Of the 217 travelers, 3 (1.4%) developed fever during the 14-day quarantine period and were hospitalized after a telehealth assessment.

Of the 28 quarantined travelers evaluated with telehealth, one traveler with diabetes presented with dizziness and weakness on the fifth day and was suspected to be hypoglycemic during the telehealth clinical evaluation. The patient's blood glucose test indicated a level of 75 mg/dl when health care workers wearing PPE visited this individual.

Clinical Features of Hospitalized Patients

Figure 2 shows the progression of clinical symptoms in three hospitalized travelers. One traveler developed fever and illness 2 days after quarantine, and two additional travelers developed fever 3 days after being admitted to the centralized quarantine center. Chest x-rays of all three patients were negative for pneumonia. Moreover, a second round of SARS-CoV-2 testing by RT-PCR throat swab was negative for COVID-19 in all three individuals. They were subsequently discharged from the hospital on April 5 and returned to the Yangmingshan quarantine center to complete the mandatory quarantine.

Figure 2. Time course of clinical symptoms in the three hospitalized travelers.



Cost Analysis of the 14-Day Mandatory Quarantine of Taiwanese Travelers

Table 2 shows the labor and overall cost incurred during the 14-day mandatory quarantine. The major personnel consisted of police officers (n=18), followed by the MDT (n=15). Moreover, the major expense was the labor cost of the MDT (US $72,334, 37.30%), followed by police officers (US $66,427, 34.25%) and the cost of PPE (US $16,016, 8.26%). The total cost incurred during the 14-day quarantine was US $193,938, which equated to US $894 per traveler.

Table 2. Cost analysis of the 14-day mandatory quarantine of 217 Taiwanese travelers.

Variable	Personnel, n	Cost (US $), amount (%)
Personnel costs		
Multidisciplinary medical team	15	72,334 (37.30)
Police officers	18	66,427 (34.25)
Janitors	3	10,880 (5.61)
Logistic group	10	8033 (4.14)
Administration staff	1	894 (0.46)
Nonpersonnel costs		
Telehealth equipment	N/A[a]	2838 (1.46)
Personal protective equipment	N/A	16,016 (8.26)
Disinfecting equipment	N/A	115 (0.06)
Infectious waste disposal	N/A	1400 (0.72)
Staff uniform disinfection	N/A	6667 (3.44)
Meals and daily supplies for quarantined travelers	N/A	8334 (4.30)
Total	47	193,938 (100)

[a] N/A: not applicable.

Discussion

Principal Findings

This cohort study reports a model for the screening and quarantining of international travelers who visited countries with a declared COVID-19 outbreak. All travelers tested negative for COVID-19 during the 14-day mandatory quarantine period. Three travelers with fever were hospitalized after the telehealth assessment, and subsequent tests for COVID-19 were negative for all three patients. The total costs during the quarantine were US $193,938, which equated to US $894 per individual. Our study demonstrates that strict infectious control measures, proactive screening, and an MDT integrated with the use of telehealth contributed to the successful quarantine of international travelers while remaining cost-effective for containing the spread of COVID-19.

Quarantine is an effective strategy to control and prevent COVID-19 outbreaks [9]. However, monitoring individuals in quarantine is essential to provide prompt management and early detection of COVID-19 cases. By establishing a line of communication between health care workers and quarantined individuals, telehealth can provide timely assessment of quarantined individuals and fast-track the hospitalization of people who develop symptoms of COVID-19. Although telehealth has not been widely adopted in monitoring COVID-19 infections in quarantined individuals, one study in China [15] used a telehealth system to monitor 188 home-quarantined individuals; it was found that 74 individuals (39.4%) were infected with SARS-CoV-2. Moreover, 6 of the 74 confirmed COVID-19 cases (8.1%) were hospitalized after the telehealth assessment [15]. Our study used telehealth to monitor COVID-19 infections in 217 international travelers during the 14-day mandatory quarantine. Three travelers with fever were hospitalized after the telehealth assessment, and subsequent tests for COVID-19 were negative for all three patients. Because telehealth can assist in providing timely assessment and does not delay hospitalization of quarantined individuals, our study suggests that it is imperative to adopt telehealth to monitor COVID-19 infections in this population.

Our report is among the first to analyze the cost-effectiveness of an enforced quarantine program. We found that the total cost of a 14-day quarantine for 217 international travelers was US $193,938, equating to US $894 per traveler. Although none of the 217 Taiwanese travelers tested positive for SARS-CoV-2 during the 14-day mandatory quarantine period, this study highlights the importance of screening and immediately quarantining returning international travelers because of the high risk of COVID-19 in this population [16]. The quarantine strategy in our report is also corroborated by a recent study [14], in which it was reported that the quarantine strategy is efficient in curbing the COVID-19 outbreak, and its relative efficacy increases when supplemented with other measures designed to reduce disease transmission.

LINE-based telehealth can provide timely assessment and care for patients who require it. During the COVID-19 surge, LINE-based telehealth could provide care for nonsevere COVID-19 cases in hospitals [13], and health care workers can provide timely assessment and management of patients with COVID-19 through the telehealth system and reduce the transmission of SARS-CoV-2 in health care facilities [11].

Limitations

The present study has two limitations. First, our study did not compare the cost-effectiveness of a 14-day quarantine for high-risk COVID-19 individuals. In this cohort study, none of the 217 Taiwanese travelers tested positive for SARS-CoV-2 during the 14-day mandatory quarantine period. As SARS-CoV-2 is highly contagious [17], it is imperative to screen and quarantine travelers who have visited countries with a declared COVID-19 outbreak. Second, the external validity of our findings may be a concern because all our patients were Taiwanese. The generalizability of our results to other non-Asian ethnic groups requires further verification. However, our findings suggest new avenues for future research.

Conclusion

This prospective cohort study reports an enforced quarantine program integrated with a telehealth system to monitor COVID-19 infection in quarantined individuals. Telehealth not only provided timely management of quarantined individuals with illness but also reduced the risk of COVID-19 infection in health care workers. Our study suggests that it is imperative to screen and quarantine international returning travelers to reduce the nationwide spread of COVID-19.

Acknowledgments

The authors are grateful to the members of the Research Office for Health Data, Department of Education and Research, Taipei City Hospital, Taiwan, for their valuable contributions to the data management and statistical analysis.

Authors' Contributions

YFY, YFT, VYS, SYC, WRY, HH, CMH, LCW, and SJH substantially contributed to the conception and design of the study, data analysis, data interpretation, and drafting of the manuscript. YFY, YFT, VYS, SYC, HH, CMH, CCC, LCW, and SJH substantially contributed to data acquisition and interpretation of the results.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Symptoms and telehealth management of 28 quarantined travelers.
[DOCX File , 14 KB - jmir_v22i12e22703_app1.docx]

References

1. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis 2020 Jul;26(7):1470-1477. [doi: 10.3201/eid2607.200282] [Medline: 32255761]
2. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/ [accessed 2020-11-13]
3. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Apr 14;323(14):1341-1342. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
4. Cheng H, Li S, Yang C. Initial rapid and proactive response for the COVID-19 outbreak - Taiwan's experience. J Formos Med Assoc 2020 Apr;119(4):771-773 [FREE Full text] [doi: 10.1016/j.jfma.2020.03.007] [Medline: 32222336]
5. Prevention and Control of COVID-19 in Taiwan. Taiwan Centers for Disease Control and Prevention. URL: https://www.cdc.gov.tw/En/Category/Page/0vq8rsAob_9HCi5GQ5jH1Q [accessed 2020-11-13]
6. Surveillance of COVID-19. Taiwan Centers for Disease Control and Prevention. URL: https://www.cdc.gov.tw/En [accessed 2020-11-13]
7. Ahn D, Shin H, Kim M, Lee S, Kim H, Myoung J, et al. Current Status of Epidemiology, Diagnosis, Therapeutics, and Vaccines for Novel Coronavirus Disease 2019 (COVID-19). J Microbiol Biotechnol 2020 Mar 28;30(3):313-324 [FREE Full text] [doi: 10.4014/jmb.2003.03011] [Medline: 32238757]
8. Rabby MII. Current Drugs with Potential for Treatment of COVID-19: A Literature Review. J Pharm Pharm Sci 2020 Apr 04;23(1):58-64. [doi: 10.18433/jpps31002] [Medline: 32251618]
9. Wilder-Smith A, Freedman D. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J Travel Med 2020 Mar 13;27(2):32052841 [FREE Full text] [doi: 10.1093/jtm/taaa020] [Medline: 32052841]
10. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. Cochrane Database Syst Rev 2020 Apr 08;4:CD013574. [doi: 10.1002/14651858.CD013574] [Medline: 32267544]
11. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
12. Lin C, Tseng W, Wu J, Tay J, Cheng M, Ong H, et al. A Double Triage and Telemedicine Protocol to Optimize Infection Control in an Emergency Department in Taiwan During the COVID-19 Pandemic: Retrospective Feasibility Study. J Med Internet Res 2020 Jun 23;22(6):e20586 [FREE Full text] [doi: 10.2196/20586] [Medline: 32544072]
13. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
14. Eilersen A, Sneppen K. Cost-benefit of limited isolation and testing in COVID-19 mitigation. Sci Rep 2020 Oct 29;10(1):18543. [doi: 10.1038/s41598-020-75640-2] [Medline: 33122753]
15. Xu H, Huang S, Qiu C, Liu S, Deng J, Jiao B, et al. Monitoring and Management of Home-Quarantined Patients With COVID-19 Using a WeChat-Based Telemedicine System: Retrospective Cohort Study. J Med Internet Res 2020 Jul 02;22(7):e19514 [FREE Full text] [doi: 10.2196/19514] [Medline: 32568727]
16. Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Ichii H, Zacharski M, et al. The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak. J Microbiol Immunol Infect 2020 Jun;53(3):467-472 [FREE Full text] [doi: 10.1016/j.jmii.2020.03.026] [Medline: 32299783]
17. Wang L, Wang Y, Ye D, Liu Q. Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence. Int J Antimicrob Agents 2020 Jun;55(6):105948 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105948] [Medline: 32201353]

Abbreviations

PPE: personal protective equipment
RT-PCR: reverse transcriptase–polymerase chain reaction
MDT: multidisciplinary team

Edited by G Eysenbach; submitted 21.07.20; peer-reviewed by E Franco, KC Wong; comments to author 06.11.20; revised version received 17.11.20; accepted 30.11.20; published 11.12.20.

Please cite as:
Yen YF, Tsai YF, Su VYF, Chan SY, Yu WR, Ho H, Hou CM, Chen CC, Woung LC, Huang SJ
Use and Cost-Effectiveness of a Telehealth Service at a Centralized COVID-19 Quarantine Center in Taiwan: Cohort Study
J Med Internet Res 2020;22(12):e22703
URL: http://www.jmir.org/2020/12/e22703/
doi:10.2196/22703
PMID:33259324

Original Paper

Research Output and International Cooperation Among Countries During the COVID-19 Pandemic: Scientometric Analysis

Nadja Grammes[1*], MSc; Dominic Millenaar[2*], MD; Tobias Fehlmann[1], MSc; Fabian Kern[1], MSc; Michael Böhm[2], MD; Felix Mahfoud[2*], MD; Andreas Keller[1*], PhD

[1]Saarland University, Chair for Clinical Bioinformatics, Saarbrücken, Germany
[2]University Hospital of Saarland, Department of Internal Medicine III, Homburg, Germany
[*]these authors contributed equally

Corresponding Author:
Dominic Millenaar, MD
University Hospital of Saarland
Department of Internal Medicine III
Kirrberger Str
Homburg, 66421
Germany
Phone: 49 68411615000
Email: dominic.millenaar@uks.eu

Abstract

Background: The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has instigated immediate and massive worldwide research efforts. Rapid publication of research data may be desirable but also carries the risk of quality loss.

Objective: This analysis aimed to correlate the severity of the COVID-19 outbreak with its related scientific output per country.

Methods: All articles related to the COVID-19 pandemic were retrieved from Web of Science and analyzed using the web application SciPE (science performance evaluation), allowing for large data scientometric analyses of the global geographical distribution of scientific output.

Results: A total of 7185 publications, including 2592 articles, 2091 editorial materials, 2528 early access papers, 1479 letters, 633 reviews, and other contributions were extracted. The top 3 countries involved in COVID-19 research were the United States, China, and Italy. The confirmed COVID-19 cases or deaths per region correlated with scientific research output. The United States was most active in terms of collaborative efforts, sharing a significant amount of manuscript authorships with the United Kingdom, China, and Italy. The United States was China's most frequent collaborative partner, followed by the United Kingdom.

Conclusions: The COVID-19 research landscape is rapidly developing and is driven by countries with a generally strong prepandemic research output but is also significantly affected by countries with a high prevalence of COVID-19 cases. Our findings indicate that the United States is leading international collaborative efforts.

(J Med Internet Res 2020;22(12):e24514) doi:10.2196/24514

KEYWORDS

scientometric analysis; COVID-19; SARS-CoV-2; citation analysis; research; literature; citation

Introduction

The global pandemic caused by the novel coronavirus SARS-CoV-2, leading to the disease COVID-19, has instigated immediate and massive worldwide research activities. Literature on preprint servers is increasing enormously. Prominent servers such as bioRxiv and medRxiv receive numerous new manuscripts each day and currently list 6063 articles (as of July 4, 2020). Additionally, peer-reviewed literature is growing at an unprecedented rate with articles published in various leading medical and related journals [1-3]. Rapid publication of research data can be desirable but also carries the risk of quality loss. In fact, some manuscripts have been accepted on the day of submission, which calls into question the completion of a sufficient peer-review process [4], leading to a relatively high number of retractions even in high-ranking journals [5,6].

This scientometric study aimed at providing profound insights into the current scientific SARS-CoV-2 research landscape. According to the World Health Organization, on July 4, 2020, the United States reported the highest absolute number of confirmed COVID-19 cases with 2,724,433 positive test results and 128,481 associated deaths [7]. In Europe, the United Kingdom and Italy reported the highest number of infected persons, with 284,280 and 241,184 cumulative cases, respectively, while China reported 85,287 cases [7]. The present study also aimed to correlate the severity of the COVID-19 outbreak with COVID-19–related scientific output per region during the pandemic, as well as to assess international collaboration.

Methods

Data Search Strategy

The online database Web of Science Core Collection (WoS) was searched to retrieve all analyzed data, containing the words “covid19,” “covid-19,” “sarscov2,” or “sars-cov-2” in the title or abstract. We refrained from adding the word “corona” to our search term, as this may identify publications unrelated to the COVID-19 pandemic. The exact search term in WoS was as follows: [TI=(covid19 OR covid-19 OR sarscov2 OR sars-cov-2) OR AB=(covid19 OR covid-19 OR sarscov2 OR sars-cov-2)]. All articles found through this search were eligible and analyzed up until and including the date of retrieval on June 14, 2020. A second search was performed on October 25 to assess changes in the scientific landscape following the initial search. There were no exclusion criteria if the article was identified by the above-mentioned search terms, including no restrictions on language, article type, or region of publication. A cross-check was performed with other medical databases such as PubMed to avoid missing articles.

Data Acquisition and Processing

By applying the web application SciPE (science performance evaluation), a dedicated web-based scientometric tool, the full set of research items was analyzed, as described elsewhere [8].

In brief, metadata of the retrieved publication data extracted from WoS were processed and visualized accordingly. WoS is the standard database for citation analyses, as it provides more details compared with other medical databases [9]. Hence, SciPE was programmed to process WOS metadata for further analysis. All institution-specific data were compared to a normalized and comprehensive list of an online university ranking list [10]. All data were coupled to a fee-based Google API (application programming interface) key enabling the assessment of exact geo-positions for all analyzed institutions by internal processing utilizing SciPE. Consecutively, institution heatmaps were created according to these results. Information on each country’s population size was extracted from the World Factbook [11].

Assessment of Collaborations Between Institutions and Countries

To assess the level of collaboration between institutions and different countries, the affiliations of the first author were analyzed and compared with the affiliations of all other coauthors. Each institution of a country that was distinct from the first author’s country was counted as one cooperation and visualized in a chord diagram. The width of each chord is proportional to the amount of existing cooperation between institutions or countries.

Ethical Approval

Since this was a metadata analysis of published work, ethics committee approval was not required.

Results

Overview

In the initial search on July 4, 2020, a total of 7185 publications were extracted from WoS, including 2592 articles, 2091 editorial materials, 2528 early access papers (likely comprising mostly of articles and letters), 1479 letters, 633 reviews, and other contributions (Figure 1). Of note, some publications could be classified into various categories (Figure 2). For example, 1014 publications fell in the categories article and early access at the same time. An additional 670 items were both letters and early access papers; interestingly, 15 items were categorized as early access and correction. Of all articles, 0.8% (58/7185) were corrections or retractions of published material. As of October 25, a total of 44,944 articles were identified on WoS with the same search terms used in the initial search. Of these, 21,218 (47.2%) were original articles, 8727 (19.4%) were editorial material, 8389 (18.7%) were letters, 4634 (10.3%) were reviews, and 342 (0.8%) were corrections or retractions.

Figure 1. Publications included in the study. Bubble size reflects the number of instances of each item class.

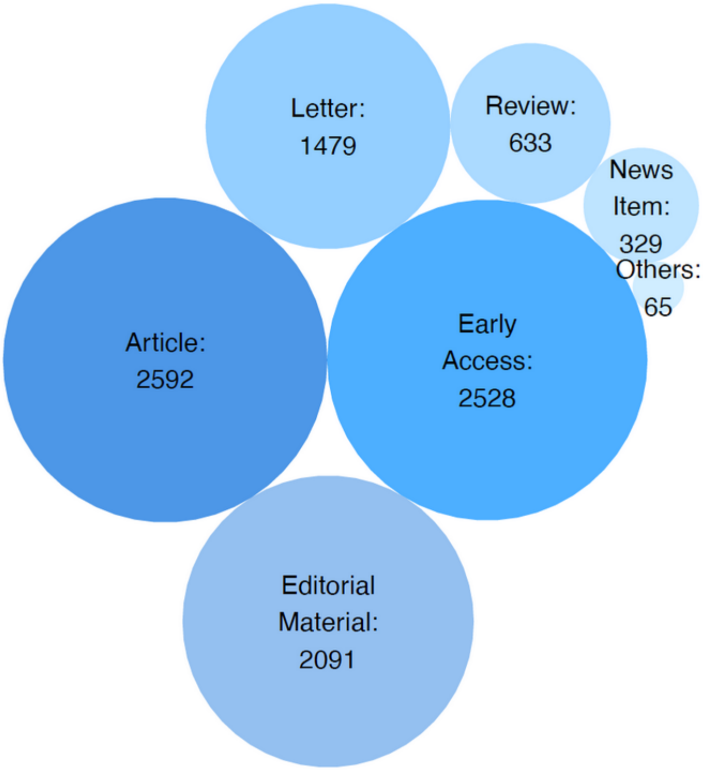
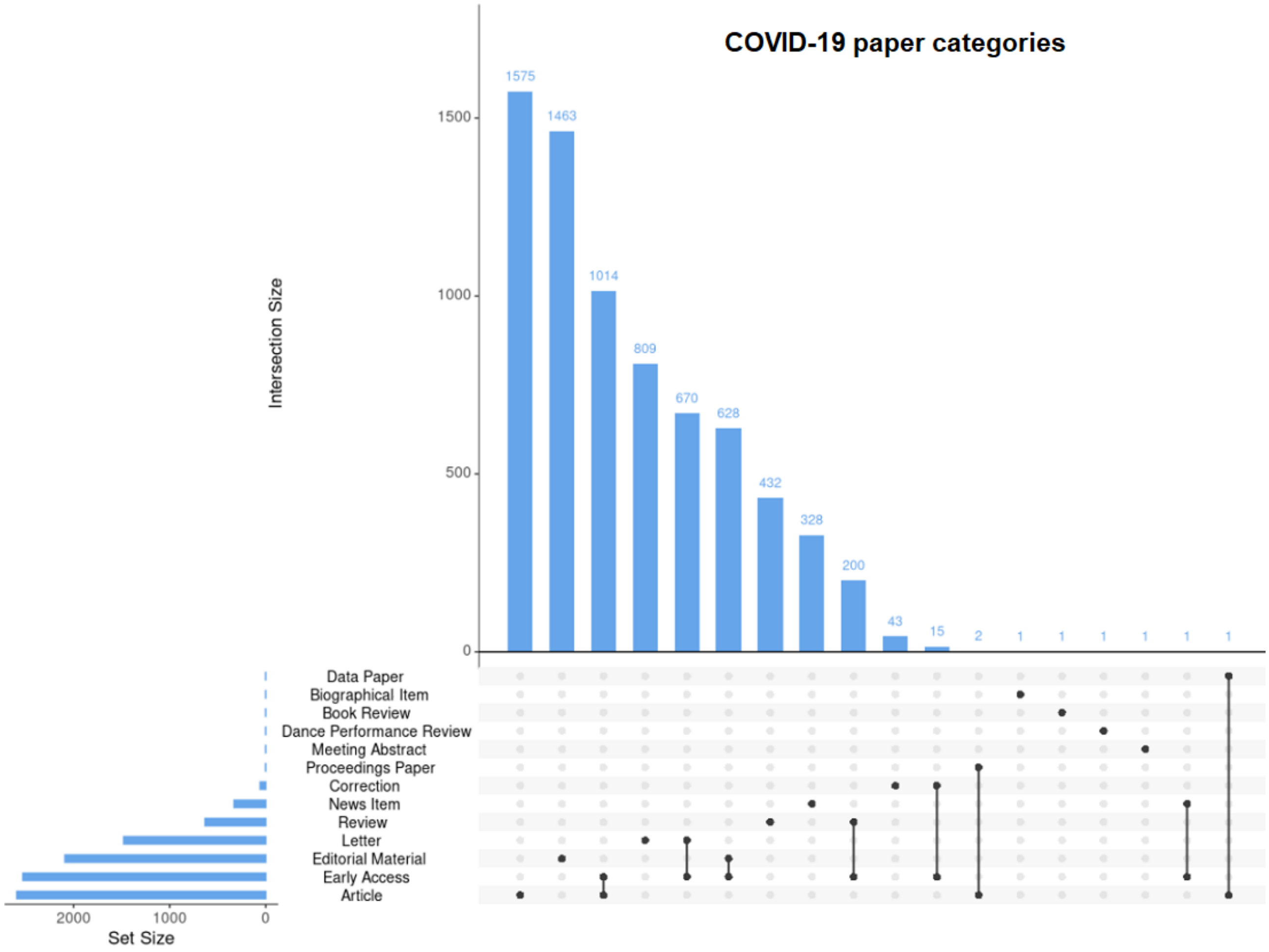


Figure 2. UpSet plot that provides details on publications that are attributed to more than one category.



The top 3 authors were medical journalists not affiliated with research or academic institutions, who published news updates. The fourth most active author (with 25 senior authorship positions) primarily published "letters to the editor," commenting on various medical research fields.

International Collaboration in COVID-19 Research

The analyses of collaborative literature on SARS-CoV-2 and COVID-19 revealed a significant amount of joint publications (Figure 3A). The cooperation landscape for the leading countries (United States, United Kingdom, China, and Italy) are highlighted in thumbnail graphics (Figure 3B). For China, the United States was the most common cooperation partner, followed by the United Kingdom. Italy also shared several manuscripts with the United States and the United Kingdom. Here, China played a far less significant role while neighboring European countries such as France, Germany, and Switzerland were frequently found to collaborate. Similarly, researchers from other European countries such as Spain often coauthored publications with researchers from Italy. The international publication behavior did not significantly change between July and October 2020: the United States, the United Kingdom, China, and Italy remained the leading nations in terms of the number of publications. The extent of international collaboration has been stable (Multimedia Appendices 1-4). With respect to universities and institutions, the Chinese University of Hong Kong played a leading role (Figure 4), collaborating on research with many different institutions, both in China and other countries. Likewise, the Wuhan University and the Massachusetts College of Pharmacy and Health Sciences shared several national and international publications. Of note, frequently only one or two manuscripts between the respective universities were found.

Figure 3. (A) Research collaborations identified by joint publications. Each edge corresponds to a joint publication between the connected countries. (B) Leading countries—United States, United Kingdom, China, and Italy—in the cooperation landscape.

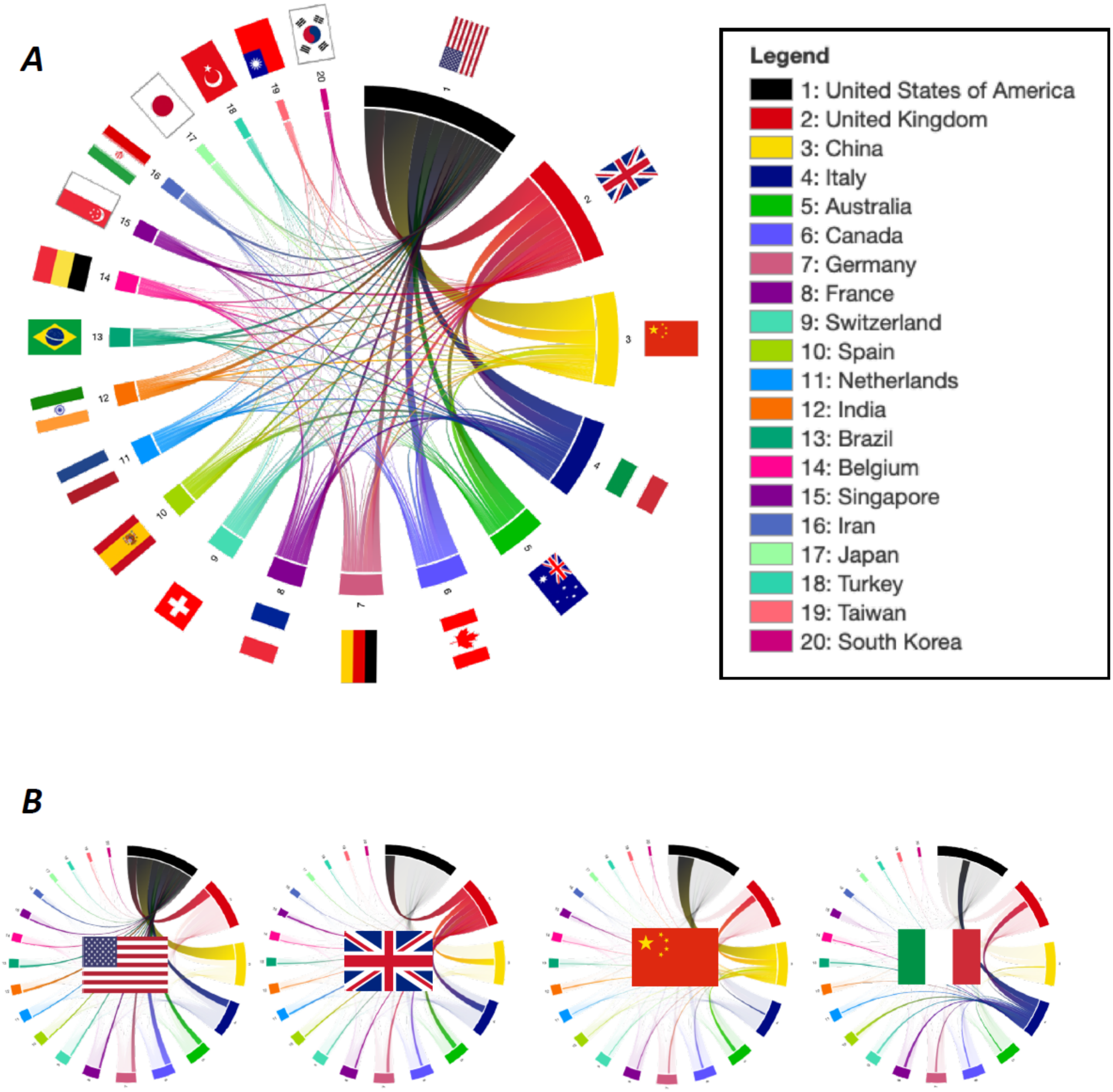
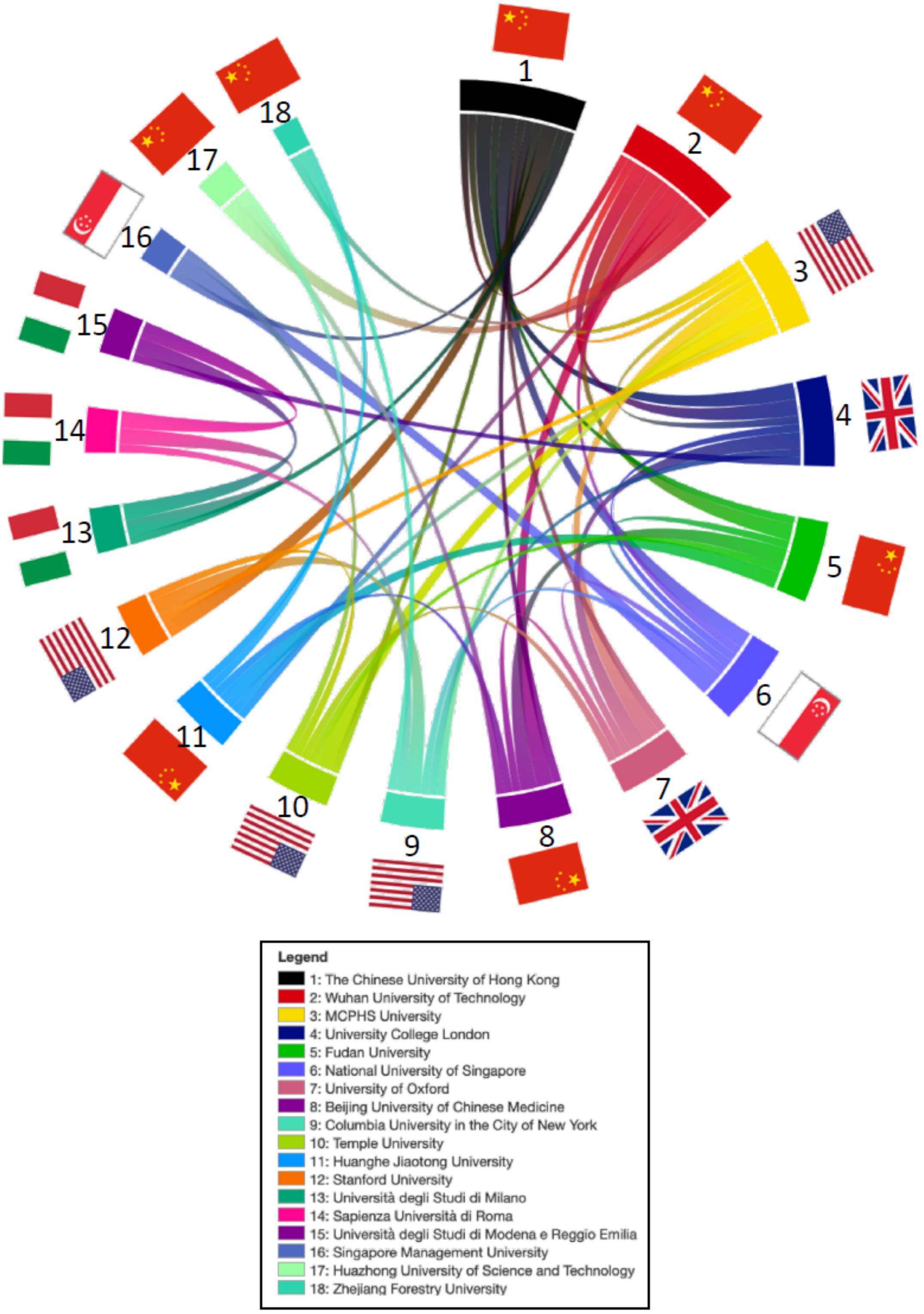


Figure 4. COVID-19 research collaboration according to institutions. Each edge connects 2 institutions that share a publication. The flags in the middle represent the countries where the institutions are located. MCPHS: Massachusetts College of Pharmacy and Health Sciences.



Research Topics

The majority of research items were published in the topic “general and internal; medicine” (Table 1), followed by “environmental and occupational health; public,” “nuclear medicine and medical imaging; radiology,” “infectious diseases,” “surgery,” “otorhinolaryngology; surgery,” and “dermatology.” The topic “virology” was found at position 8.

Table 1. Research categories of publications matched to their respective topics.

Topics	Publications, n (%)
General and internal; medicine	836 (11.64)
Environmental and occupational health; public	291 (4.0)
Nuclear medicine and medical imaging; radiology	194 (2.70)
Infectious diseases	171 (2.38)
Surgery	149 (2.07)
Otorhinolaryngology; surgery	146 (2.03)
Dermatology	143 (1.99)
Virology	139 (1.93)
Pharmacology and pharmacy	138 (1.92)
Cardiac and cardiovascular systems	123 (1.71)
Psychiatry	121 (1.68)
Anesthesiology	103 (1.43)
Critical care medicine	101 (1.41)
Oncology	98 (1.36)
Ophthalmology	91 (1.27)

Regional Differences in COVID-19 Research

With 1806 research items, the United States was the leading country in terms of COVID-19–related publications, followed by China (n=1306), Italy (n=856), and the United Kingdom (n=817). Spain and France, both of which were seriously affected European countries, were in positions 9 and 11, respectively. Focusing on the first or last author, the patterns were similar and the ranking of the top countries remained unchanged. In terms of continents, Europe, North America, and Asia published a similar number of research items (Figure 5). According to publications in relation to confirmed COVID-19 cases or related deaths and total population size, the United States had the highest number of both COVID-19 cases and related publications. In Europe, Italy was one of the leading countries. On the other hand, China had a lower number of cases and fewer deaths compared with the United States, although the population count was higher (Figure 6). Multimedia Appendices 5-8 visualize institution heatmaps for 4 highly affected countries, analyzing their research output per research institution. In Italy, Milano and Bergamo accounted for the highest number of COVID-19–related publications (n=131), followed by Rome (n=81) and Padua (n=78), as indicated by the numbers in the heatmap in Multimedia Appendix 5. Another highly represented region was Naples. In China, Wuhan led in terms of research volume, followed by Beijing, Shanghai, and Hong Kong/Shenzhen (Multimedia Appendix 6). In the United States, the east coast (Boston, New York, and Philadelphia) and California (Los Angeles and the San Francisco Bay area) had the highest output (Multimedia Appendix 7). Here, areas with a generally strong prepandemic research output that were highly affected by COVID-19 contributed the most amount of publications pertaining to the pandemic. Likewise, Detroit and Chicago were highly represented in the publication statistics. In France, Paris had the strongest research output while other French regions had no relevant manuscript numbers published to date (Multimedia Appendix 8).

Figure 5. Country coverage of publications. The donut chart represents continents (AS: Asia; EU: Europe; NA: North America; SA: South America; OC: Oceania; AF: Africa) and countries contributing to the publication output.

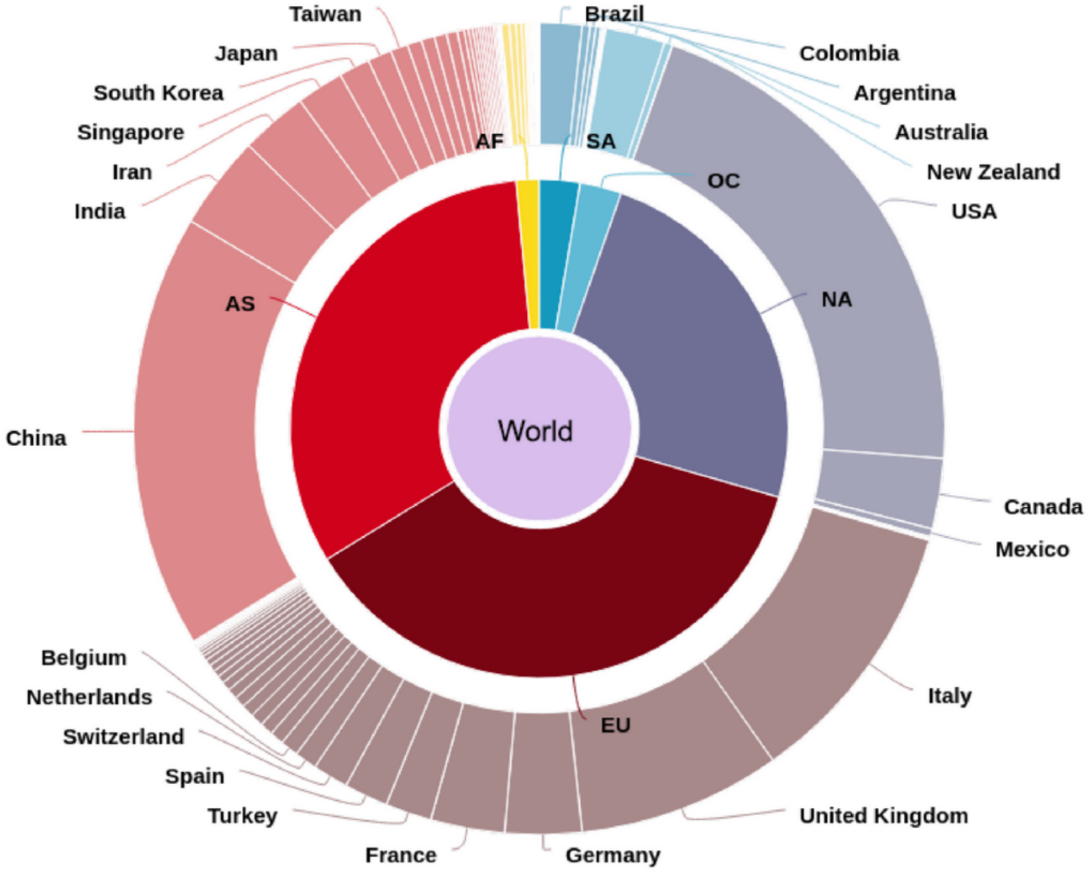
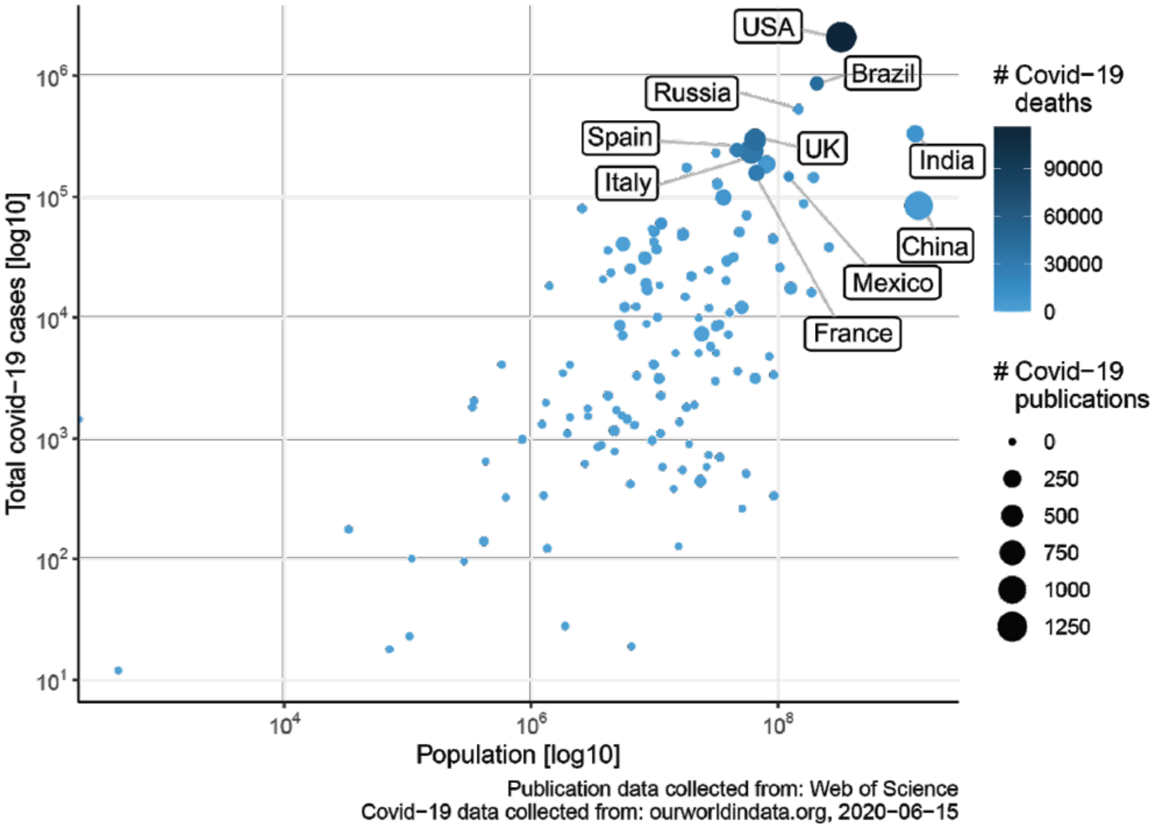


Figure 6. Population of countries versus the number of reported COVID-19 cases. Bubble sizes represent the number of publications; the color of the bubble represents the number of deaths. Data obtained from ourworldindata.org [12].



Publication data collected from: Web of Science
Covid−19 data collected from: ourworldindata.org, 2020−06−15

The leading 10 institutions were Wuhan University of Technology (Wuhan, China), Università degli Studi di Milano (Milano, Italy), Massachusetts College of Pharmacy and Health Sciences University (Boston, United States), The Chinese University of Hong Kong (Hong Kong, China), Fudan University (Shanghai, China), Columbia University in the City of New York (New York, United States), National University of Singapore (Singapore), Singapore Management University (Singapore), University of Oxford (Oxford, United Kingdom), and Ankara University (Ankara, Turkey).

Discussion

Principal Findings

This scientometric analysis provided profound insights into the publication landscape of COVID-19 research during the first months this disease was declared a pandemic. Countries severely affected by the pandemic such as Italy as well as those with a generally high research output such as the United States contributed significantly to the literature base. There were several retractions of published articles, indicating questionable peer-review processes and flawed data integrity. International collaborations were extensive, especially in countries with high numbers of COVID-19 cases, with an obvious underrepresentation of cooperation between China and Italy.

Considering the most active authors of COVID-19–related articles, some specifics must be acknowledged. The publications by the three most active authors were not original research articles but consisted mainly of letters. Interestingly, these journalistic articles providing news updates were indexed in WoS, PubMed, and other scientific databases along with other research work. As these articles play an important role in the visualization of the scientific landscape on COVID-19, these articles were included in our analysis for a more comprehensive picture. A recent study on coronavirus-related research in general revealed that medical journals have sped up their publication and production process during the pandemic. Indeed, the turnover time was reduced by 49% from submission to acceptance, which was mainly driven by a decrease in peer-review time [13]. One may speculate that this expedited review process was related to publication pressure by researchers submitting papers but also by journals aiming to publish articles with high-citation likelihood, which could enhance the relative importance of a journal within its field [4]. In line, a relatively large proportion of retractions and corrections of COVID-19–related articles was identified herein, adding up to 0.8% of all published materials, and was also found in high-impact journals [14,15]. A recent study on coronavirus research in the last decade found a large proportion of open access articles. From 2001-2020, 59.2% of all research articles on coronavirus research were provided free of charge. This number significantly increased in 2020 to 91.4%, mostly related to research on COVID-19 [16]. This high percentage of open access to scientific information and open data is crucial to facilitate better and faster research toward a vaccine and inform public health measures essential to contain the spread of the virus.

Interestingly, most articles were published in the topic “general and internal; medicine” as opposed to the topic “virology,” ranking at position 8. This may be driven by the overall higher aggregate impact factor (IF) in the category “general and internal; medicine” (aggregate IF 4.386) compared with “virology” (aggregate IF 3.731) [17], making the first mentioned journal category a generally more attractive option for article submissions.

Among the countries involved in COVID-19–related research, Italy played an exceptional role. Given the population and the general research output of Italy, it is disproportionately represented. Mapping the number of COVID-19 cases or COVID-19–associated deaths to countries, it becomes clear that Italy, as one of the most severely hit countries in Europe, also showed the largest scientific output. It is worth mentioning that the regions that suffered most from the pandemic (such as Milano, Bergamo, Bologna, and Padua) had an exceptional research output. In other countries, such as France, the majority of publications originated from the capital, whereas regions highly affected by the pandemic in peripheral areas were underrepresented.

Our analysis revealed a wide global collaboration network between several publishing countries. Here, the United States was a leading collaborator, sharing a significant amount of manuscript authorships with the United Kingdom, China, and Italy. Collaborations in medical research have been seen in other medical fields before [18]. This cooperation is often found between neighboring countries but also, as in the present case, between countries interested in similar research areas. It becomes obvious, however, that China and Italy, despite both being highly affected as well as productive in terms of research efforts, lack collaboration with one another. This analysis focused on the first wave of the COVID-19 pandemic; however, in a recent second analysis, the scientific landscape in this area including international collaborations remained similar.

Limitations

It is important to mention some limitations of our scientometric study. We rely on input from WoS, which is dependent on input query. Efforts were made to include as many specific publications as possible while simultaneously avoiding false positives, along with performing cross-checks with other medical databases to ensure a comprehensive data analysis. This scientometric analysis is of a quantitative, not qualitative, nature. One measure to assess the research quality would be to analyze citations, but given the comparably short time in which thousands of manuscripts have been published, a more comprehensive analysis can be expected in the future. If citation numbers grow, this will allow further analyses, according to three easily interpretable parameters: productivity, total impact, and how successful an author has been so far, as proposed in a recent study [19]. Regarding the number of COVID-19 cases and related deaths, we relied on published data from official authorities. However, this depends on both the integrity of these self-selected numbers as well as the extent of diagnostic testing in each country. Herein, we focused on COVID-19–associated death rates, as other variables such as the number of cases and hospitalizations provide only rough estimates based on the case fatality rate [20].

Conclusion

The publication landscape of COVID-19 is rapidly developing, making it challenging to identify high-quality research that substantially adds to the current knowledge base. Almost 1% of the literature considered in this study were corrections or retractions of articles, which challenges the quality and integrity of the expedited review process. The high number of publications is driven by countries with a generally strong research output in the past, but this also includes countries heavily affected by the pandemic such as Italy. In terms of international cooperation, the United States is most active while China is underrepresented. The most obvious finding is an underrepresentation of joint publications between China and Italy, despite both being strongly affected by the COVID-19 pandemic and producing a high research output.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (SFB TRR219) to MB and FM and by the Deutsche Gesellschaft für Kardiologie to MB and FM. The funders had no role in the design of this study, its execution, analyses, interpretation of the data, or decision to submit results for publication.

Authors' Contributions

NG and DM made substantial contributions to the conception of the work; both authors analyzed and interpreted the data, and drafted the manuscript. TF and FK made substantial contributions to data acquisition and revised the manuscript critically for important intellectual content. MB made substantial contributions to the concept of the work and revised the manuscript critically for important intellectual content. FM and AK made substantial contributions to the conception and design of the work as well as the interpretation of the data; both revised the manuscript critically for important intellectual content. All authors gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

MB is supported by Deutsche Gesellschaft für Kardiologie (DGK) and Deutsche Forschungsgemeinschaft (SFB TRR219) and has received speaker honoraria from Abott, Amgen, Astra Zeneca, Bayer, Boehringer Ingelheim, Bristol Myers Squibb, Medtronic, Novartis, Servier and Vifor. FM is supported by Deutsche Gesellschaft für Kardiologie (DGK), and Deutsche

Forschungsgemeinschaft (SFB TRR219) and has received scientific support and speaker honoraria from Bayer, Boehringer Ingelheim, Medtronic and ReCor Medical. The remaining authors declare no conflict of interest.

Multimedia Appendix 1
Heatmap representing Italy. The numbers denote the publication count in the corresponding region and are color-coded (search date: October 25, 2020).
[PNG File , 891 KB - jmir_v22i12e24514_app1.png]

Multimedia Appendix 2
Heatmap representing China. The numbers denote the publication count in the corresponding region and are color-coded (search date: October 25, 2020).
[PNG File , 1155 KB - jmir_v22i12e24514_app2.png]

Multimedia Appendix 3
Heatmap representing the United States. The numbers denote the publication count in the corresponding region and are color-coded (search date: October 25, 2020).
[PNG File , 1483 KB - jmir_v22i12e24514_app3.png]

Multimedia Appendix 4
Heatmap representing France. The numbers denote the publication count in the corresponding region and are color-coded (search date: October 25, 2020).
[PNG File , 1584 KB - jmir_v22i12e24514_app4.png]

Multimedia Appendix 5
Heatmap representing Italy. The numbers denote the publication count in the corresponding region and are color-coded (search date: June 14, 2020).
[PNG File , 642 KB - jmir_v22i12e24514_app5.png]

Multimedia Appendix 6
Heatmap representing China. The numbers denote the publication count in the corresponding region and are color-coded (search date: June 14, 2020).
[PNG File , 1005 KB - jmir_v22i12e24514_app6.png]

Multimedia Appendix 7
Heatmap representing the United States. The numbers denote the publication count in the corresponding region and are color-coded (search date: June 14, 2020).
[PNG File , 527 KB - jmir_v22i12e24514_app7.png]

Multimedia Appendix 8
Heatmap representing France. The numbers denote the publication count in the corresponding region and are color-coded (search date: June 14, 2020).
[PNG File , 478 KB - jmir_v22i12e24514_app8.png]

References

1. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
2. Steensels D, Oris E, Coninx L, Nuyens D, Delforge M, Vermeersch P, et al. Hospital-Wide SARS-CoV-2 Antibody Screening in 3056 Staff in a Tertiary Center in Belgium. JAMA 2020 Jul 14;324(2):195-197. [doi: 10.1001/jama.2020.11160] [Medline: 32539107]
3. Sakurai A, Sasaki T, Kato S, Hayashi M, Tsuzuki S, Ishihara T, et al. Natural History of Asymptomatic SARS-CoV-2 Infection. N Engl J Med 2020 Aug 27;383(9):885-886 [FREE Full text] [doi: 10.1056/NEJMc2013020] [Medline: 32530584]
4. Armstrong S. Research on covid-19 is suffering. BMJ 2020 May 28;369:m2045. [doi: 10.1136/bmj.m2045] [Medline: 32467100]
5. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Retraction: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. N Engl J Med. DOI: 10.1056/NEJMoa2007621. N Engl J Med 2020 Jun 25;382(26):2582 [FREE Full text] [doi: 10.1056/NEJMc2021225] [Medline: 32501665]

6. Mehra MR, Ruschitzka F, Patel AN. Retraction—Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. The Lancet 2020 Jun 13;395(10240):1820 [FREE Full text] [doi: 10.1016/S0140-6736(20)31324-6] [Medline: 32511943]
7. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-07-04]
8. Scholz SS, Dillmann M, Flohr A, Backes C, Fehlmann T, Millenaar D, et al. Contemporary scientometric analyses using a novel web application: the science performance evaluation (SciPE) approach. Clin Res Cardiol 2020 Jul;109(7):810-818. [doi: 10.1007/s00392-019-01568-x] [Medline: 31686209]
9. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2008 Feb;22(2):338-342. [doi: 10.1096/fj.07-9492LSF] [Medline: 17884971]
10. Universities Search Engine. uniRank. URL: https://www.4icu.org/ [accessed 2020-06-14]
11. The World Factbook. Central Intelligence Agency. 2020. URL: https://www.cia.gov/library/publications/the-world-factbook/ [accessed 2020-07-04]
12. Our World in Data. URL: https://ourworldindata.org/ [accessed 2020-07-04]
13. Horbach S. Pandemic Publishing: Medical journals drastically speed up their publication process for Covid-19. bioRxiv. Preprint posted online April 18, 2020. [doi: 10.1101/2020.04.18.045963]
14. Wu J, Leung K, Bushman M, Kishore N, Niehus R, de SP. Addendumtimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine. Preprint posted online March 13, 2020. [doi: 10.21203/rs.3.rs-17453/v1]
15. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. The Lancet Infectious Diseases 2020 May;20(5):553-558. [doi: 10.1016/s1473-3099(20)30144-4]
16. Belli S, Mugnaini R, Baltà J, Abadal E. Coronavirus mapping in scientific publications: When science advances rapidly and collectively, is access to this knowledge open to society? Scientometrics 2020 Jul 01:1-25 [FREE Full text] [doi: 10.1007/s11192-020-03590-7] [Medline: 32836526]
17. Journal Impact Factors. Journal Citation Reports. Clarivate Analytics. 2020. URL: https://clarivate.com/webofsciencegroup/solutions/journal-citation-reports/ [accessed 2020-07-04]
18. Millenaar D, Fehlmann T, Scholz S, Pavlicek V, Flohr A, Dillmann M, et al. Research in Atrial Fibrillation: A Scientometric Analysis Using the Novel Web Application SciPE. JACC Clin Electrophysiol 2020 Aug;6(8):1008-1018. [doi: 10.1016/j.jacep.2020.05.010] [Medline: 32819514]
19. Siudem G, Żogała-Siudem B, Cena A, Gagolewski M. Three dimensions of scientific impact. Proc Natl Acad Sci U S A 2020 Jun 23;117(25):13896-13900. [doi: 10.1073/pnas.2001064117] [Medline: 32513724]
20. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases 2020 Jun 30;20(6):669-677 [FREE Full text] [doi: 10.1016/S1473-3099(20)30243-7] [Medline: 32240634]

Abbreviations

API: application programming interface
IF: impact factor
SciPE: science performance evaluation
WoS: Web of Science

Edited by G Eysenbach; submitted 22.09.20; peer-reviewed by J Zhang, L Suppan; comments to author 16.10.20; revised version received 04.11.20; accepted 09.11.20; published 11.12.20.

Please cite as:
Grammes N, Millenaar D, Fehlmann T, Kern F, Böhm M, Mahfoud F, Keller A
Research Output and International Cooperation Among Countries During the COVID-19 Pandemic: Scientometric Analysis
J Med Internet Res 2020;22(12):e24514
URL: http://www.jmir.org/2020/12/e24514/
doi:10.2196/24514
PMID:33237878

Viewpoint

Addressing Public Health Emergencies via Facebook Surveys: Advantages, Challenges, and Practical Considerations

André Grow[1], PhD; Daniela Perrotta[1], PhD; Emanuele Del Fava[1], PhD; Jorge Cimentada[1], PhD; Francesco Rampazzo[2], MSc; Sofia Gil-Clavel[1], MSc; Emilio Zagheni[1], PhD

[1]Laboratory of Digital and Computational Demography, Max Planck Institute for Demographic Research, Rostock, Germany
[2]Saïd Business School and Leverhulme Centre for Demographic Science, University of Oxford, Oxford, United Kingdom

Corresponding Author:
André Grow, PhD
Laboratory of Digital and Computational Demography
Max Planck Institute for Demographic Research
Konrad-Zuse-Str 1
Rostock, 18057
Germany
Phone: 49 3812081142
Email: grow@demogr.mpg.de

Abstract

Surveys of the general population can provide crucial information for designing effective nonpharmaceutical interventions to tackle public health emergencies, such as the COVID-19 pandemic. Yet, conducting such surveys can be difficult, especially when timely data collection is required. In this viewpoint paper, we discuss our experiences with using targeted Facebook advertising campaigns to address these difficulties in relation to the COVID-19 pandemic. We describe central advantages, challenges, and practical considerations. This includes a discussion of potential sources of bias and how they can be addressed.

(J Med Internet Res 2020;22(12):e20653) doi:10.2196/20653

KEYWORDS

Facebook; web-based surveys; public health emergency; COVID-19

Introduction

As of September 9, 2020, the COVID-19 pandemic has caused over 27.4 million cases and over 894,000 deaths around the world [1]. To control the spread of COVID-19, national and local governments have implemented nonpharmaceutical interventions, including school closures, bans on large gatherings, mobility restrictions, and physical isolation, as well as unprecedented measures like local and nationwide lockdowns. Such measures have likely been critical in delaying and containing the COVID-19 pandemic [2]. However, individual behaviors, rather than governmental actions, may be crucial for controlling the spread of COVID-19 in the long run [3]. Timely and accurate data on human behaviors are thus of paramount importance in closely monitoring the adoption of preventive measures, emergence of symptoms, and changes in mobility and person-to-person contacts in the population.

In this context, surveys of the general population can provide central information needed to assess people's acceptance of and compliance with behavioral guidelines. Such surveys are also needed to capture spontaneous bottom-up behavioral changes. Yet, researchers who want to conduct surveys that directly address ongoing epidemics are faced with unique methodological challenges, as follows: (1) these surveys need to be designed, implemented, and conducted quickly, as epidemics spread rapidly and are difficult to predict, especially when they involve new emerging diseases (ie, timeliness); (2) they need to cover the entire population, and in the event of large-scale epidemics or pandemics, they need to be conducted simultaneously in multiple countries or regions, as regional differences could be relevant for designing effective interventions (ie, coverage); and (3) they should be cost-effective, as obtaining large research funds quickly for an ad hoc survey can be difficult (ie, cost-effectiveness). Furthermore, in the case of COVID-19, the nature of recommended social distancing measures may limit some traditional modes of data collection, such as face-to-face interviews and even phone interviews, to the extent that they rely on large call centers.

In this viewpoint, we discuss the use of Facebook as a recruitment tool to address these challenges. Our assessment is based on the COVID-19 Health Behavior Survey (CHBS) that we conducted between March 13 and August 12, 2020 in 8 countries (ie, Belgium, France, Germany, Italy, the Netherlands, Spain, the United Kingdom, and the United States). Participant recruitment took place on a daily basis via targeted advertisements on Facebook, resulting in a total of 144,034 completed questionnaires. In what follows, we first provide an overview of the most important design aspects of the survey, and then discuss some of the central advantages, challenges, and practical considerations related to using Facebook advertisements in surveys that address public health emergencies, such as the COVID-19 pandemic. This includes a discussion and empirical assessment of potential sources of bias and how they can be addressed. Additionally, we make recommendations for those who want to implement similar surveys in the near future, and we hope that this will facilitate timely data collection to address the current—and possibly future—public health and societal crises.

Methodological Approach

The CHBS is a web-based survey that focuses on people's reactions to the COVID-19 pandemic and targets individuals aged ≥18 years. The questionnaire has 4 sections, which encompass sociodemographic characteristics (eg, age, sex, and education), health indicators (eg, underlying medical conditions), behaviors and attitudes related to COVID-19 (eg, perceived threat level and preventive measures taken), and social contacts (ie, the number of interactions with other people).

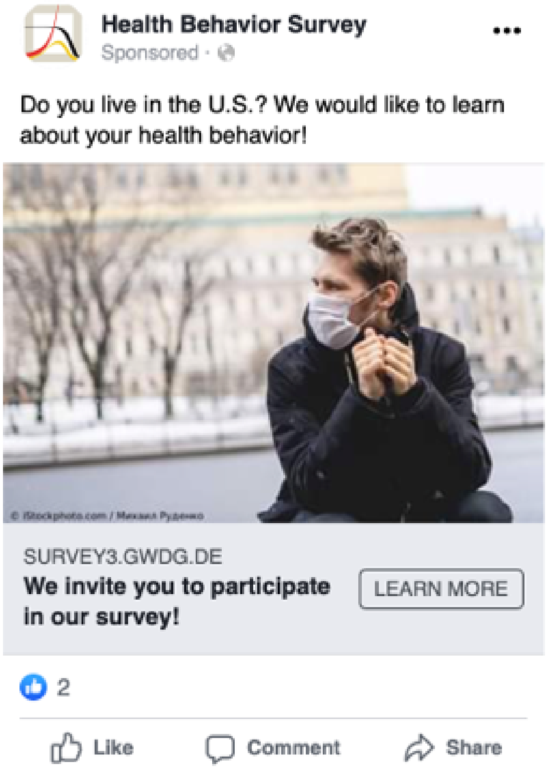
Recruitment took place via Facebook, by means of advertisement campaigns that we created with the Facebook Ads Manager (FAM). The Facebook business model centers on targeted advertisements, and the FAM enables advertisers to create campaigns that can be directed at specific user groups. Targeting can be based on both users' demographic characteristics and a set of characteristics that Facebook infers from users' behavior on the social network. Advertising campaigns consist of 3 levels. The highest level is the campaign level, at which the goals of the campaign are defined (eg, generating awareness or generating traffic). The second level is the ad set level, at which the target audience, budget, and schedule are defined. The third level includes the advertisements themselves, which can consist of multiple advertising materials (eg, images and videos), advertising text, and the link to the page where Facebook users should be directed when they click on the ad. More details on these levels can be found in the Pötzschke and Braun [4] study.

We created 1 campaign per country and stratified each campaign at the ad set level by users' sex (ie, male and female), age group (ie, 18-24, 25-44, 45-64, and ≥65 years), and region of residence. In the European countries, the region classification largely followed the NUTS (nomenclature of territorial units for statistics)-1 classification, which we aggregated into larger macro regions. In the United States, the region classification was based on census regions. More details on region stratification can be found in the Perrotta et al [5] study. Our study design resulted in 24-56 strata per country. Each ad set contained 6 different images, leading to a total of 1776 different ads. Figure 1 provides an example of the ads shown to Facebook users in the United States. We launched the campaigns between March 13, 2020 (ie, in Italy, the United Kingdom, and the United States) and April 4, 2020 (ie, in Belgium). This difference in the timing of the inclusion of countries is owed to a trade-off between the time needed to translate and technically implement country-specific surveys and the goal to start data collection in a timely manner. We concluded the survey on August 12, 2020.

Facebook advertising campaigns can have different goals, and the overall costs that are incurred will partly depend on the chosen goal. We chose the goal of generating traffic. This meant that the Facebook algorithms would optimize ad delivery to maximize the likelihood of people clicking on the ad when it is shown to them. Advertisers can choose and define how Facebook should use their budget to meet these goals. For example, advertisers can set a budget that is evenly spread over a fixed period, or they can define an average daily budget that Facebook would seek to meet weekly over an unspecified period. We opted for the latter, as the duration of the COVID-19 pandemic was uncertain. Based on these parameters, ad delivery was determined through an automated bidding process, in which a given ad competes for delivery with ads from other advertisers who are targeting the same user groups. In this process, Facebook considers the budget that can be afforded for delivering a given ad and the likelihood of the ad being of interest to users by comparing it to competing advertisements. Before an ad campaign is launched, the FAM provides an estimate of various parameters, such as the size of the target audience and daily reach, which makes it possible to gauge the likely performance of the advertising campaign.

It is important to note that our study is not the first to use targeted Facebook advertisements for participant recruitment in health research. Earlier research has used this approach to address topics such as smoking behavior [6], cannabis use [7], and mental health [8]. Furthermore, Whitaker et al [9] and Thornton et al [10] have performed systematic reviews of related literature. However, compared to these studies, our survey stands out because of its cross-national character, duration, and population coverage, given that we continuously collected data from 8 countries for 5 months and recruited more than 140,000 participants from most—if not all—parts of society and subnational regions.

Figure 1. Example of an advertisement in the Facebook advertising campaign in the United States.



Do you live in the U.S.? We would like to learn about your health behavior!

We invite you to participate in our survey!

Advantages

Our use of Facebook for participant recruitment enabled us to address the challenges of timeliness, coverage, and cost-effectiveness.

With regard to the advantages of our recruitment methodology, first, we were able to design, implement, and launch the survey in a timely manner. To summarize, preparing a Facebook advertising campaign involves creating an advertising account, a Facebook page that is associated with the advertisements/survey, and the ads themselves. Ad creation can be performed in bulk, by uploading a CSV (comma-separated values) file. Once the ads have been created, they need to be submitted for review, during which their compliance with Facebook advertising policies is assessed. This review can take between a couple of hours and a day, or longer. In our case, it was usually completed within 24 hours. However, it should be noted that in April/May 2020, Facebook warned advertisers that reviews could be delayed due to the COVID-19 pandemic [11]. Once reviewed, the ads can be delivered. Ads only need to be reviewed again when major changes are made (eg, changes to the advertising materials or ad text).

The second advantage is that our use of Facebook enabled us to draw multinational samples from diverse parts of the respective national populations. Facebook is the largest social media platform, with 2.45 billion monthly active users worldwide as of Fall 2019 [12]. In the United States, about 68% of adults used Facebook in 2018, whereas this percentage is 56% in Germany, 75% in Italy and Spain, 76% in France, 79% in the Netherlands, 85% in the United Kingdom, 89% in Belgium [13,14]. Compared to younger and middle-aged adults, older adults generally tend to be underrepresented on Facebook [15]. Nevertheless, the Facebook user population provides a cross section of the overall population with access to the internet. A comparison of different social media platforms (ie, Facebook, LinkedIn, Twitter, Tumblr, and Reddit) in the United States has suggested that Facebook is the most representative in terms of users' educational attainment and internet skills [16].

The third advantage is that our use of Facebook made participant recruitment comparatively cost-effective, even though it can be difficult to determine the exact costs in advance. This is partly due to the nature of the bidding process that determines ad delivery and variation in the competition for advertising space. A central performance measure of Facebook advertising campaigns is the cost per click (CPC) value. Whitaker et al [9] performed a review of studies that used Facebook for recruiting participants in health research, and they reported CPC values between €0.17 (US $0.20) and €1.46 (US $1.74). This variation is likely due to differences in the definitions of targeted user groups, the competition from other advertisers, and the likelihood of users clicking on the respective ad. Our costs were similar to those reported in earlier research. Between March 13 and August 12, 2020, we collected 144,034 questionnaires at an overall CPC of about €0.14 (US $0.17) and an overall cost per completed questionnaire (CPCQ) of about €1.05 (US $1.25), excluding the value-added tax. It should be noted that as a relief measure, the value-added tax in Germany was temporarily reduced from 19% to 16% on July 1, 2020. The difference between the CPC and CPCQ is due to Facebook users who clicked on one of our ads, but did not complete our questionnaire.

Thanks to these advantages, we were able to collect data that provided key insights into attitudes and behaviors that shape—and are shaped by—the COVID-19 pandemic. Figure

2 illustrates this by plotting the average number of face-to-face social contacts that respondents reported for the day before participating in the survey for the entire observation period. Face-to-face social contacts are the main vehicle for virus spread, given that the SARS-CoV-2 virus is mainly transmitted by infected secretions or respiratory droplets [17]. Figure 2 shows that there was great variation in the number of face-to-face contacts over time, especially in those that occurred outside the home. When paired with external information on the COVID-19 pandemic, such as lockdown measures and infection rates, these data provide valuable insights into the effectiveness of different policies. When the data is further broken down by respondents' demographic characteristics, it also becomes possible to assess whether different demographic groups respond differently to different policies. The data can also be used to calculate central epidemiological metrics (eg, the effective reproduction number, Rt) and design more realistic epidemiological models. More details on the contact patterns that we observed in the CHBS can be found in the Del Fava et al [18] study. Insights into other behaviors and attitudes toward the COVID-19 pandemic can be found in the Perrotta et al [5] study.

Figure 2. Average number of contacts at home and outside of home per week between March 13 and August 12, 2020 across the 8 countries. Lines show averages and shaded areas show 95% confidence intervals. Poststratification weighting has been applied.

To better illustrate the comparative strengths of our approach, it is helpful to contrast our work with similar ad hoc web-based survey efforts that also focus on COVID-19. A prominent example is the work of Fetzer et al [19], who used an open survey to recruit 108,075 participants from 58 countries between March 20 and April 7, 2020. In this study, recruitment took place via link sharing on social media and similar channels. Another prominent example is the work of De Coninck et al [20], who used an existing opt-in online panel that was maintained by a commercial polling company to recruit 1000 participants in the Flemish region of Belgium between March 17 and March 22, 2020. Both studies used web-based surveys to collect information about people's attitudes and behaviors toward COVID-19. Compared to Fetzer et al [19], our use of Facebook offered more control over the recruitment process, given that when a survey link is shared via social media, it is not possible to control who is invited to participate in the survey. Hence, our targeted advertising methodology made it arguably easier to ensure that our samples were demographically balanced. Compared to De Coninck et al [20], our use of Facebook arguably offered less control over the recruitment process, given that existing online panels typically offer more detailed information about prospective participants than the FAM. These panels make it easier to collect demographically balanced samples. However, our use of Facebook offered a larger reach in terms of the number of countries that could be included and the time frame that was covered. In terms of costs, the CPCQ that our paid advertisements incurred was higher than the CPCQ incurred by Fetzer et al [19], as their approach to link sharing did not incur any costs. However, based on our personal communication with De Coninck et al [20], the CPCQ we incurred was similar to that of De Coninck et al [20], who paid a commercial polling company for data collection.

Challenges

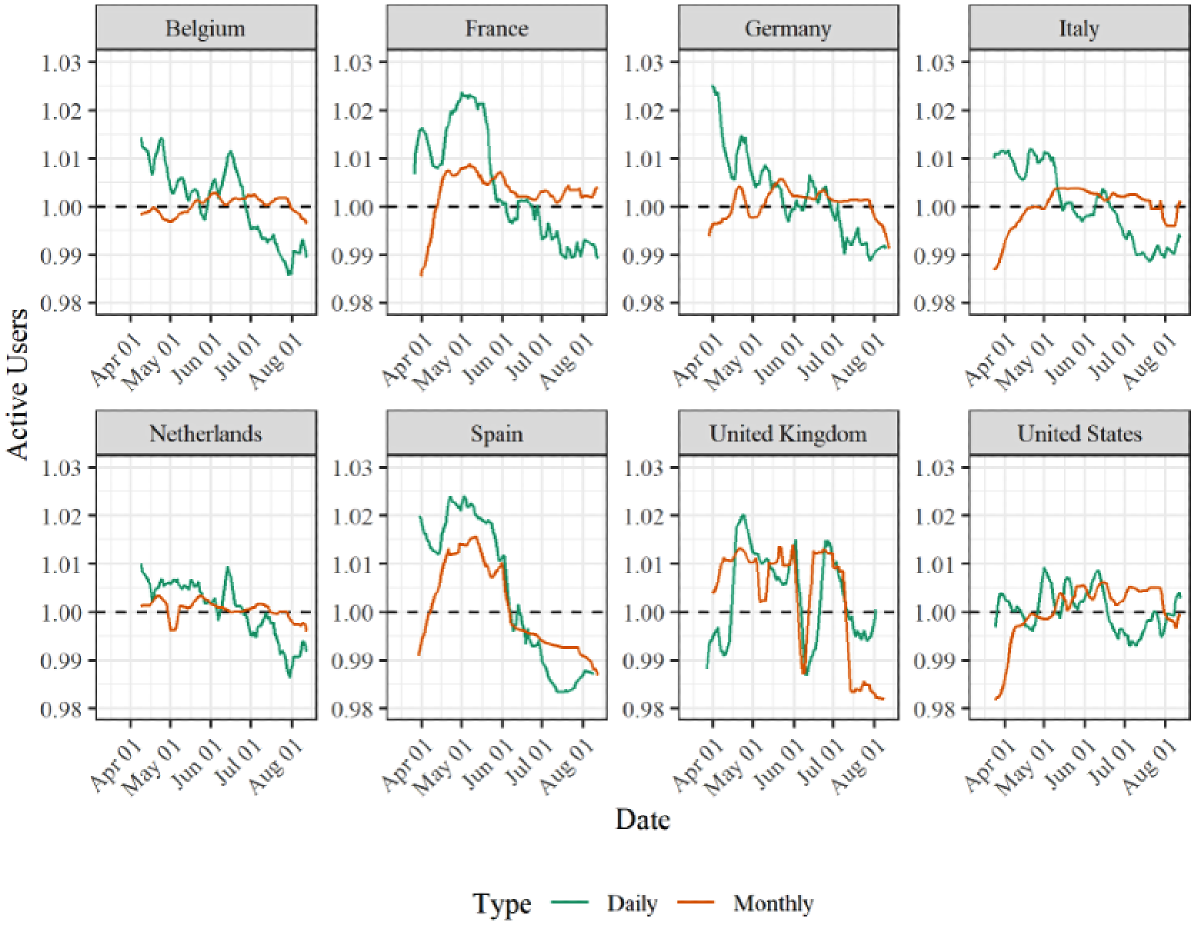
Eliciting information via self-administered web-based surveys involves several challenges. For example, issues with recall inaccuracy often occur when some time has elapsed between a specific event of interest and participation in the survey [21,22]. This puts limits on the information that can be collected, and such issues should therefore be considered in light of the goals of the respective study. If, for example, detailed and accurate medical information is essential (eg, exact blood pressure measurements), a web-based survey may not be the best choice, and an in-person assessment with medically trained personnel may be preferable. Discussing the methodological challenges of web-based surveys is out of the scope of this paper. Instead, we focus on the challenges that are specific to using Facebook for participant recruitment. A discussion on the methodological challenges of web-based surveys can be found in the Eysenbach and Wyatt [23] study.

The most important set of challenges relates to the issue of self-selection bias. The Facebook user base is a rough cross section of the overall population with internet access, but not all demographic groups are equally well represented [24]. Additionally, there may be variation in Facebook users' interest in the survey topic. Hence, there is no guarantee that the resulting samples will be representative in terms of central demographic characteristics (eg, age and sex) and important unobservable characteristics. This issue is potentially exacerbated by the algorithmic optimization that Facebook uses for ad delivery. If certain demographic groups are more likely to click on an ad than others, Facebook might increasingly deliver the ads to these groups, thereby reinforcing existing self-selection bias. This is particularly difficult to correct if survey participation and ad delivery are affected by user characteristics that cannot be easily considered when defining the relevant sampling strata.

It is important to note that if a survey is conducted over a long period of time, there may be changes in Facebook user activity. This may be due to seasonal variation in people's use of Facebook [25]. However, it seems possible that the development of the COVID-19 pandemic may have also led to changes in the composition of our samples over time. In the early days of the pandemic especially, the SARS-CoV-2 virus dominated the news, and lockdown measures were put in place to curb its spread. This may have increased participation in our survey in 2 ways. First, a lack of alternative activities due to lockdown measures may have led people from various sub-populations to spend more time on Facebook than normal, and this may have increased the likelihood of seeing our ads. Second, the salience of the pandemic may have increased the chance of people clicking on our ads when seeing them. Over time, as the number of infections decreased and lockdown measures eased, participation in the survey may have decreased, and the resulting samples may have become more selective.

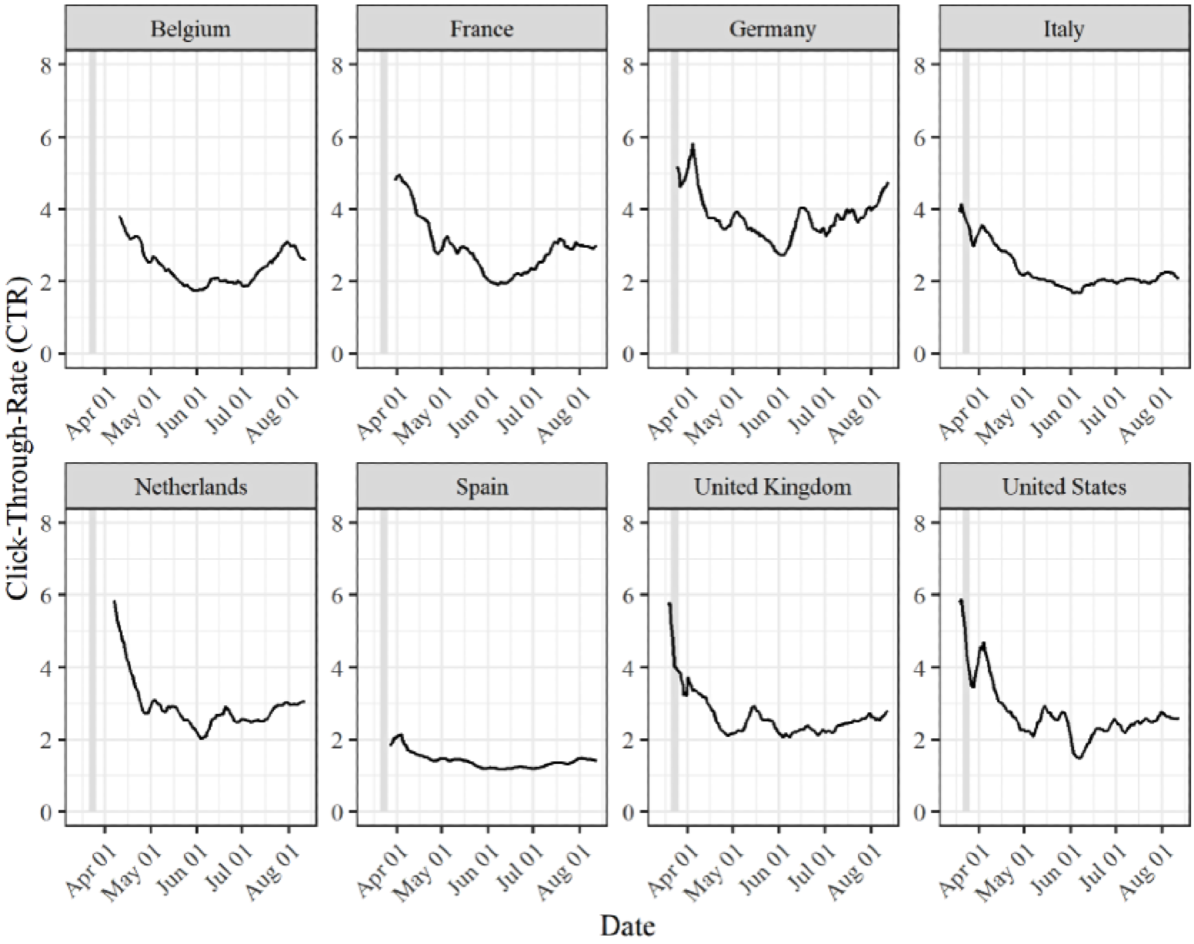
Figures 3 and 4 show changes in user behavior and survey participation over time. This was done by plotting the average number of daily active Facebook users (DAUs) and monthly active Facebook users (MAUs), as well as the click-through rates (CTRs), for all countries over the entire study period. All 3 measures are based on estimates from Facebook, which we obtained via the Facebook application programming interface. The number of DAUs is the number of unique active users on a given day, whereas the number of MAUs provides the number of unique users who have been active on Facebook within the last 30 days [26]. Both estimates are commonly used to assess the potential reach of advertising campaigns, and we systematically collected DAU and MAU estimates for all our strata over the entire study period. The CTR is defined as the number of people who click on an ad after seeing it, and CTRs become available after a campaign has been started and delivered to users [27]. Hence, changes in the number of DAUs and MAUs provide insights into changes in Facebook use, whereas changes in the CTR provide insights into topic salience/interest among Facebook users. It should be noted that between March 21 and March 26, 2020, we experienced technical problems with ad delivery across several countries, leading to a substantially lower number of participants than in the other weeks of our study. The CRT values for this period are therefore less reliable than those for the rest of the survey period.

Figure 3. Number of DAUs and MAUs between March 13 and August 12, 2020 across the 8 countries based on all ad sets. Numbers were standardized for each country by dividing the value for a given day by the average number of DAUs/MAUs over the entire observation period. Lines show 7-day moving averages. We collected DAU and MAU values every 6 hours, as these estimates can change within 1 day. We averaged these multiple observations to obtain 1 number per day. Data collection started between March 18 and April 4; no data was collected on April 11, April 12, and between August 3 and 6 due to technical issues. DAU: daily active Facebook user; MAU: monthly active Facebook user.



As shown in Figure 3, the number of DAUs was largest during the early weeks of the observation period for most countries. However, this number gradually decreased, usually by 3-5 percentage points. This means that the number of unique users who could have seen our ads on a given day decreased over time. The only exceptions to this were the United Kingdom and the United States, where changes in the number of DAUs were more erratic. The trends in the MAU values somewhat deviate from those in DAU values. Typically, there was an initial increase in the number of MAUs, but this number later decreased. Hence, while the number of individuals who may have seen our ads on a daily basis decreased over time, we may have reached users who we would not have reached in other months toward the middle of the observation period. Furthermore, in Figure 4, the CTR values show a clear trend over time. The CTR was initially high across all countries, and then it decreased before ultimately increasing again. This means that in the early phases of the survey, Facebook users were more likely to click on our ads than in later phases. Hence, our data suggest that over time, the process by which users selected themselves into the survey may have changed, but the data do not allow us to precisely assess why these changes occurred.

Figure 4. CTRs between March 13 and August 12, 2020 across the 8 countries. The line shows the 7-day moving average. The shaded area shows the period during which technical problems with ad delivery occurred. CTR: click-through rate.



How can the problem of self-selection be addressed? We suggest 4 methodological steps that can help alleviate this problem. First, in line with common approaches in traditional survey research, we suggest stratifying ad campaigns based on characteristics that are known to relate to survey participation and the outcome of interest, such as in the Pötzschke and Braun [4] study. Evidently, in the case of new emerging diseases, relevant individual characteristics are difficult to know in advance. Moreover, several relevant characteristics will not be available for creating strata in the FAM (eg, preexisting medical conditions). However, characteristics that are available (eg, age and sex) should be considered for stratifying advertising campaigns. We also suggest stratifying ads by region within countries, as people's responses might vary locally. With this approach, the bias that Facebook's ad delivery algorithms may generate is counteracted, leading to more balanced samples.

The second step we suggest, which is in line with the Zagheni and Weber [28] study, is applying poststratification techniques to the samples obtained from Facebook, to the extent that they deviate from the overall population in terms of important characteristics. In this regard, using Facebook offers distinct advantages over other, less controlled ways of recruiting online samples. As indicated previously, prior to launching a campaign, the FAM provides an estimate of the size of the audience with the characteristics of interest. Arguably, this feature is similar to a sampling plan and has been used in earlier research to conduct a "virtual census" of the overall population, such as in the Zagheni et al [29] study. Furthermore, the ad performance estimates that the FAM provides after a campaign has been launched (eg, the number of users to whom a given campaign, ad set, or ad has been delivered and the number of users who have clicked on the ad) can be paired with information about survey completion rates for each stratum. This makes it possible to calculate performance measures, such as approximate participation rates. However, it is important to keep in mind that many of the measures that Facebook reports are only estimates. The resulting indicators should thus be viewed as informed proxies.

As a complement to this approach, Zhang et al [30] recently reported that by selectively activating and deactivating ad sets over the course of the survey period, it is possible to obtain representative samples from Facebook that do not require poststratification. This approach is feasible if the specific timing of participation over the study period does not matter. However, this was not feasible in our case, as our goal was to obtain daily balanced samples. Selectively closing and opening ads over the course of a day or week would have implied that responses from certain subgroups may have been concentrated during a certain time of the day or certain days of the week. As an alternative, researchers may opt for dynamically adjusting the budget, so that more money is spent on strata that are underrepresented in the survey. With this approach, it is important to keep in mind that large changes in the budget allocated to an ad set may trigger the ad review process again, which can lead to a gap in data collection. As this would have undermined the goals of our study, we decided against this approach. Instead, we continuously recruited members of all strata using a stable budget.

With regard to the third step, we suggest that the issue of self-selection due to participant characteristics that are difficult to observe before people take part in the survey can be partly addressed by considering possible sources of bias in the design of the survey and advertising campaign. In the case of our study, we expected that individuals who are particularly concerned about COVID-19 might be more likely to participate, and such concerns might also be reflected in reported behaviors and attitudes. As it is not possible to stratify Facebook advertising campaigns based on such concerns, we considered this issue in 2 ways. First, in the survey, we directly assessed participants' concerns about COVID-19 and other factors that may raise such concerns. Second, when selecting the images for our ads, we aimed to create variation in how closely the images were linked to the topic of COVID-19. In our analyses, we were able to control the extent of how all these factors affected participation and answers to other questions.

Figure 5 shows the different images that we used in the ads. We considered images 1 and 2 to be the least strongly linked to COVID-19 and images 5 and 6 to be the most strongly linked to COVID-19. In total, 141,879 of the respondents arrived at the survey via one of the ads. Of these respondents, about 74% (104,292/141,879) arrived via image 5, about 16% arrived via image 6 (23,157/141,879), about 7% (9699/141,879) arrived via image 3, and the rest arrived via the remaining images. To assess whether the image through which participants arrived at the survey was related to their concerns about COVID-19, we conducted a Kruskal-Wallis test by rank. In this test, we assessed respondents' personal perceptions of how large a threat COVID-19 presented for themselves. This was determined using a 5-point Likert-type scale, in which a score of 1 represents very low threat and a score of 5 represents very high threat. These scores were associated with the picture through which participants arrived at the survey. Table 1 shows the number of respondents who selected 4 or 5 on the scale, which indicated a high or very high threat perception, respectively. The number of respondents who perceived COVID-19 as a high or very high threat was largest among those who arrived at the survey via image 3 (7178/9280, 77%) and lowest among those who arrived via image 2 (922/1461, 63%). The observed variation in threat perceptions across images was significant at the 1% level (χ^2_5=801, $P<.01$). A broader analysis of how the different images related to respondents' self-reported attitudes and behaviors would be important, but our assessment suggests that the inclusion of different images helped with recruiting more diverse samples in terms of concerns for COVID-19.

With regard to the fourth step, in recent years, the multilevel regression and poststratification approach to making inferences from highly selected survey data [31] has proved effective in producing unbiased population estimates [32,33]. In the first stage of multilevel regression and poststratification, the sample is partitioned into a large number of demographic strata (eg, each combination of age group, sex, and region), and a multilevel regression model is used to estimate the outcome of interest, such as the average number of contacts or the percentage of people wearing a face mask, in each stratum. In the second stage of multilevel regression, the stratum-level estimates are used to produce a final population-level estimate, and poststratification weights are used to account for the proportion of each stratum in the population. This approach combined with the previously mentioned steps, and a fine partition of the sample in demographic strata enables researchers to make proper inferences at the population level, even in presence of strong selection bias.

Figure 5. Images used in the Facebook advertising campaigns.

1 – Male athlete
©Adobe Stock/grki

2 – Group of athletes
©Adobe Stock/nd3000

3 – Woman blowing nose
©iStockphoto/Goodboy Picture Company

4 – Couple blowing noses
©iStockphoto/Goodboy Picture Company

5 – Woman wearing mask
©Adobe Stock/shintartanya

6 – Man wearing mask
©iStockphoto/Михаил Руденко

Table 1. The number of respondents who personally perceived COVID-19 as a high (score=4) or very high threat (score=5) to themselves based on a 5-point Likert-type scale for each ad image. No weighting has been applied.

Image number	Image name	Total, n/N (%)
1	Male athlete	422/586 (72%)
2	Group of athletes	922/1461 (63%)
3	Woman blowing nose	7178/9280 (77%)
4	Couple blowing noses	1819/2519 (72%)
5	Woman wearing mask	69,421/102,061 (68%)
6	Man wearing mask	16,882/22,451 (75%)

Another challenge relates to trust in web-based surveys. Web-based surveys may face suspicion, as they could be used to elicit personal information for nonresearch purposes (eg, marketing, identity theft, etc) [34]. In addition, over the last several years, there have been several incidents that may have negatively affected the trust that the public has in the data protection measures put in place by Facebook. A prominent example is the Cambridge Analytica scandal, in which the personal data of Facebook users was harvested without consent, with the goal of influencing the 2016 US presidential election through microtargeting [35]. When fielding our survey, we encountered such suspicions in the commenting sections of our ads, and 1 notable concern was that Facebook would transfer personal user information to us. We addressed this issue by highlighting that the survey was anonymous and that no personal information was exchanged with Facebook. We also provided additional information about our research institute, the research team, and the goals of our survey by providing a link to our data protection policy and providing information about preliminary results and reports as they became available. Despite these measures, it seems likely that Facebook users who are concerned about data privacy were less inclined to participate in our survey. Additionally, while anonymous web-based surveys have the potential to reduce the likelihood of respondents providing socially desirable answers compared to personal interviews [36], privacy concerns may have rendered them reluctant to answer questions that they perceive as sensitive. We addressed this issue by offering the possibility of not answering questions that they feel uncomfortable with, to avoid forcing answers on sensitive topics.

Practical Considerations

There are some practical aspects that need to be considered when using Facebook ads for survey research. First, the possibility of targeting certain user groups makes it easier to recruit members from certain subpopulations, even when they are underrepresented on Facebook. Yet, it is important to keep in mind that as the number of strata in the campaign increases, so will the selectivity and costs of the campaign. This is particularly true if the members of certain strata are less likely to participate than members of other strata. For example, if the goal is to stratify a campaign based on 5-year age groups, the number of strata will increase considerably compared to strata based on 10- or 20-year age groups. This means that overall, more responses will need to be collected to have enough observations per stratum to apply poststratification weighting. Furthermore, when the members of certain groups engage little with Facebook, a larger share of the budget needs to be devoted to recruiting them.

Second, both the advertisements and the study page to which these advertisements are linked need to be actively managed. Facebook advertisements are similar to user posts, meaning that users can react to them (eg, liking), comment on them, and share them with friends. Additionally, users can leave posts on the study page and review the page itself. In our experience, it is important to engage with user comments and provide additional information if needed, to maintain trust with current and prospective participants. The time investment that this requires should not be underestimated, especially for studies that run for a long period of time and are conducted in multiple languages. On average, our ads received about 19,300 impressions and about 135 comments per day, and it took us about 1 hour to manage 50-100 comments. The exact time it took to manage comments depended on the length of the comment and the complexity of the answer that was required. Over time, we became more experienced and efficient in managing comments. Based on these numbers, researchers who want to conduct a study on a similar topic and receive a similar number of impressions per day should expect to spend about 1-2.5 hours per day managing comments.

Third, the ad review process involves an automatic assessment of the links that are provided in the ads. In this process, the webpage to which the link leads is accessed. Hence, submitting ads for review can generate a large amount of traffic for a web-based survey, and it is important to keep in mind that the number of times that the survey page was accessed is not equivalent to the number of potential respondents who have accessed it, as some of the traffic may have been generated by the review process. It would therefore not be valid to approximate the survey completion rate by dividing the number of completed questionnaires by the number of page accesses. Furthermore, it is important to schedule sufficient time between submitting ads for review and launching the data collection. In our case, the review process was usually completed in a timely manner. However, there might be delays (eg, those predicted by Facebook due to COVID-19) and problems (eg, rejections due to violations of the Facebook advertising policies) that increase the time between review completion and delivery.

Conclusion

To conclude, we suggest that targeted advertisements on Facebook can be a powerful tool for recruiting participants in ad hoc surveys of the general population during a public health emergency, as long as certain methodological steps are taken to address the issue of self-selection. We hope that the experiences that we have described here, together with our recommendations, will make it easier for other researchers to implement similar surveys that tackle current and future pandemics.

Acknowledgments

We thank the editor and an anonymous reviewer for providing helpful comments on an earlier version of this paper. FR conducted most of his research for this article as a PhD student at the Department of Social Statistics and Demography of the University of Southampton with a scholarship from the Economic and Social Research Council (South Coast Doctoral Training Partnership; Project: ES/P000673/1).

Conflicts of Interest

None declared.

References

1. Weekly Operational Update on COVID-19, 9 September 2020. World Health Organization. URL: https://www.who.int/docs/default-source/coronaviruse/weekly-updates/wou-9-september-2020-cleared-14092020.pdf?sfvrsn=68120013_2 [accessed 2020-12-08]
2. Ferguson N, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team.

2020 Mar 16. URL: https://spiral.imperial.ac.uk/bitstream/10044/1/77482/14/2020-03-16-COVID19-Report-9.pdf [accessed 2020-12-08]
3. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020 Mar 21;395(10228):931-934 [FREE Full text] [doi: 10.1016/S0140-6736(20)30567-5] [Medline: 32164834]
4. Pötzschke S, Braun M. Migrant Sampling Using Facebook Advertisements. Soc Sci Comput Rev 2016 Sep 08;35(5):633-653 [FREE Full text] [doi: 10.1177/0894439316666262]
5. Perrotta D, Grow A, Rampazzo F, Cimentada J, Del Fava E, Gil-Clavel S, et al. Behaviors and attitudes in response to the COVID-19 pandemic: Insights from a cross-national Facebook survey. medRxiv. Preprint published online on July 15, 2020. [FREE Full text] [doi: 10.1101/2020.05.09.20096388]
6. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012 Feb 23;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
7. Borodovsky JT, Marsch LA, Budney AJ. Studying Cannabis Use Behaviors With Facebook and Web Surveys: Methods and Insights. JMIR Public Health Surveill 2018 May 02;4(2):e48 [FREE Full text] [doi: 10.2196/publichealth.9408] [Medline: 29720366]
8. Choi SH, Milne DN, Glozier N, Peters D, Harvey SB, Calvo RA. Using different Facebook advertisements to recruit men for an online mental health study: Engagement and selection bias. Internet Interv 2017 Jun;8(1):27-34 [FREE Full text] [doi: 10.1016/j.invent.2017.02.002] [Medline: 30135825]
9. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
10. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
11. Delays Expected for Requested Ad Reviews | Facebook Business Help Center. Facebook Inc. URL: https://www.facebook.com/business/help/285854632398488 [accessed 2020-04-15]
12. Facebook Reports Third Quarter 2019 Results. Facebook Inc. 2019 Oct 30. URL: https://investor.fb.com/investor-news/press-release-details/2019/Facebook-Reports-Third-Quarter-2019-Results/default.aspx [accessed 2020-04-15]
13. Social Media Use 2018: Demographics and Statistics. Pew Research Center. 2018 Mar 01. URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2020-04-15]
14. Gendronneau C, Wisniowski A, Yildiz D, Zagheni E, Fiorio L, Hsiao Y, et al. Measuring labour mobility and migration using big data: Exploring the potential of social-media data for measuring EU mobility flows and stocks of EU movers. Employment, Social Affairs & Inclusion: Publications Office of the European Union 2019 [FREE Full text] [doi: 10.2767/474282]
15. Gil-Clavel S, Zagheni E. Demographic differentials in Facebook usage around the world. 2019 Jul 06 Presented at: Proceedings of the Thirteenth International AAAI Conference on Web and Social Media; June 11-14, 2019; Munich, Germany p. 647-650.
16. Hargittai E. Potential Biases in Big Data: Omitted Voices on Social Media. Soc Sci Comput Rev 2018 Jul 30;38(1):10-24. [doi: 10.1177/0894439318788322]
17. Transmission of SARS-CoV-2: implications for infection prevention precautions. World Health Organization. 2020 Jul 09. URL: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions [accessed 2020-09-29]
18. Del Fava E, Cimentada J, Perrotta D, Grow A, Rampazzo F, Gil-Clavel S, et al. The differential impact of physical distancing strategies on social contacts relevant for the spread of COVID-19. medRxiv Preprint posted online on May 18, 2020. [doi: 10.1101/2020.05.15.20102657]
19. Fetzer T, Witte M, Hensel L, Jachimowicz J, Haushofer J, Ivchenko A, et al. Perceptions of an Insufficient Government Response at the Onset of the COVID-19 Pandemic are Associated with Lower Mental Well-Being. PsyArXiv Preprints. Preprint posted online on April 16, 2020. [FREE Full text] [doi: 10.31234/osf.io/3kfmh]
20. De Coninck D, d'Haenens L, Matthijs K. Perceived vulnerability to disease and attitudes towards public health measures: COVID-19 in Flanders, Belgium. Pers Individ Dif 2020 Nov 01;166:110220 [FREE Full text] [doi: 10.1016/j.paid.2020.110220] [Medline: 32834279]
21. Simon GE, VonKorff M. Recall of psychiatric history in cross-sectional surveys: implications for epidemiologic research. Epidemiol Rev 1995;17(1):221-227. [doi: 10.1093/oxfordjournals.epirev.a036180] [Medline: 8521941]
22. Newell S, Girgis A, Sanson-Fisher R, Ireland M. Accuracy of patients' recall of Pap and cholesterol screening. Am J Public Health 2000 Sep;90(9):1431-1435. [doi: 10.2105/ajph.90.9.1431] [Medline: 10983202]
23. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
24. Ribeiro FN, Benevenuto F, Zagheni E. How Biased is the Population of Facebook Users? Comparing the Demographics of Facebook Users with Census Data to Generate Correction Factors. arXiv. Preprint posted online on May 16, 2020.

25. Capturing the Spirit of Summer. Facebook Inc. 2020 May 16. URL: https://www.facebook.com/business/news/insights/capturing-the-spirit-of-summer [accessed 2020-09-28]
26. Graph API Reference v9.0: Ad Campaign Delivery Estimate. Facebook Inc. URL: https://developers.facebook.com/docs/marketing-api/reference/ad-campaign-delivery-estimate/ [accessed 2020-10-09]
27. Parameters - Marketing API. Facebook Inc. URL: https://developers.facebook.com/docs/marketing-api/insights/parameters/v9.0 [accessed 2020-10-09]
28. Zagheni E, Weber I. Demographic research with non-representative internet data. Int J of Manpower 2015 Apr 07;36(1):13-25. [doi: 10.1108/ijm-12-2014-0261]
29. Zagheni E, Weber I, Gummadi K. Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants. Popul Dev Rev 2017 Oct 11;43(4):721-734. [doi: 10.1111/padr.12102]
30. Zhang B, Mildenberger M, Howe PD, Marlon J, Rosenthal SA, Leiserowitz A. Quota sampling using Facebook advertisements. Political Sci Res Methods 2018 Dec 05;8(3):558-564. [doi: 10.1017/psrm.2018.49]
31. Gelman A, Little TC. Poststratification Into Many Categories Using Hierarchical Logistic Regression (1997). Citeseer. 1997. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.5270 [accessed 2020-12-08]
32. Downes M, Gurrin LC, English DR, Pirkis J, Currier D, Spittal MJ, et al. Multilevel Regression and Poststratification: A Modeling Approach to Estimating Population Quantities From Highly Selected Survey Samples. Am J Epidemiol 2018 Aug 01;187(8):1780-1790. [doi: 10.1093/aje/kwy070] [Medline: 29635276]
33. Wang W, Rothschild D, Goel S, Gelman A. Forecasting elections with non-representative polls. Int J Forecast 2015;31(3):980-991. [doi: 10.1016/j.ijforecast.2014.06.001]
34. Evans JR, Mathur A. The value of online surveys. Internet Research 2005 Apr 01;15(2):195-219. [doi: 10.1108/10662240510590360]
35. Lapowsky I. How Cambridge Analytica Sparked the Great Privacy Awakening. Wired. 2019 Mar 17. URL: https://www.wired.com/story/cambridge-analytica-facebook-privacy-awakening/ [accessed 2020-10-09]
36. Stern MJ, Bilgen I, Dillman DA. The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design. Field methods 2014 Feb 10;26(3):284-301 [FREE Full text] [doi: 10.1177/1525822X13519561]

Abbreviations

CHBS: COVID-19 Health Behavior Survey
CPC: cost per click
CPCQ: cost per completed questionnaire
CSV: comma-separated values
CTR: click-through rate
DAU: daily active Facebook user
FAM: Facebook Ads Manager
MAU: monthly active Facebook user
NUTS: nomenclature of territorial units for statistics

Edited by G Eysenbach; submitted 25.05.20; peer-reviewed by K Hoffmann Pham, C García; comments to author 13.07.20; revised version received 12.10.20; accepted 28.10.20; published 14.12.20.

Please cite as:
Grow A, Perrotta D, Del Fava E, Cimentada J, Rampazzo F, Gil-Clavel S, Zagheni E
Addressing Public Health Emergencies via Facebook Surveys: Advantages, Challenges, and Practical Considerations
J Med Internet Res 2020;22(12):e20653
URL: http://www.jmir.org/2020/12/e20653/
doi:10.2196/20653
PMID:33284782

Original Paper

Social Media Insights Into US Mental Health During the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data

Danny Valdez[1], PhD; Marijn ten Thij[2], PhD; Krishna Bathina[2], MS; Lauren A Rutter[3], PhD; Johan Bollen[2], PhD

[1]Department of Applied Health Science, School of Public Health, Indiana University, Bloomington, IN, United States
[2]Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, United States
[3]Psychological and Brain Sciences, Indiana University, Bloomington, IN, United States

Corresponding Author:
Danny Valdez, PhD
Department of Applied Health Science
School of Public Health
Indiana University
1033 E 3rd Street
Office 143
Bloomington, IN
United States
Phone: 1 812 855 1561
Email: danvald@iu.edu

Abstract

Background: The COVID-19 pandemic led to unprecedented mitigation efforts that disrupted the daily lives of millions. Beyond the general health repercussions of the pandemic itself, these measures also present a challenge to the world’s mental health and health care systems. Considering that traditional survey methods are time-consuming and expensive, we need timely and proactive data sources to respond to the rapidly evolving effects of health policy on our population’s mental health. Many people in the United States now use social media platforms such as Twitter to express the most minute details of their daily lives and social relations. This behavior is expected to increase during the COVID-19 pandemic, rendering social media data a rich field to understand personal well-being.

Objective: This study aims to answer three research questions: (1) What themes emerge from a corpus of US tweets about COVID-19? (2) To what extent did social media use increase during the onset of the COVID-19 pandemic? and (3) Does sentiment change in response to the COVID-19 pandemic?

Methods: We analyzed 86,581,237 public domain English language US tweets collected from an open-access public repository in three steps. First, we characterized the evolution of hashtags over time using latent Dirichlet allocation (LDA) topic modeling. Second, we increased the granularity of this analysis by downloading Twitter timelines of a large cohort of individuals (n=354,738) in 20 major US cities to assess changes in social media use. Finally, using this timeline data, we examined collective shifts in public mood in relation to evolving pandemic news cycles by analyzing the average daily sentiment of all timeline tweets with the Valence Aware Dictionary and Sentiment Reasoner (VADER) tool.

Results: LDA topics generated in the early months of the data set corresponded to major COVID-19–specific events. However, as state and municipal governments began issuing stay-at-home orders, latent themes shifted toward US-related lifestyle changes rather than global pandemic-related events. Social media volume also increased significantly, peaking during stay-at-home mandates. Finally, VADER sentiment analysis scores of user timelines were initially high and stable but decreased significantly, and continuously, by late March.

Conclusions: Our findings underscore the negative effects of the pandemic on overall population sentiment. Increased use rates suggest that, for some, social media may be a coping mechanism to combat feelings of isolation related to long-term social distancing. However, in light of the documented negative effect of heavy social media use on mental health, social media may further exacerbate negative feelings in the long-term for many individuals. Thus, considering the overburdened US mental health care structure, these findings have important implications for ongoing mitigation efforts.

(J Med Internet Res 2020;22(12):e21418) doi:10.2196/21418

KEYWORDS

social media; analytics; infodemiology; infoveillance; COVID-19; United States; mental health; informatics; sentiment analysis; Twitter

Introduction

Beyond the obvious physical health ramifications of the COVID-19 pandemic, public health and the greater medical community is also bracing for a mental health crisis [1]. Within the span of 4 months, 45% of Americans indicated that the COVID-19 pandemic had taken a toll on their mental health, reporting higher levels of sadness and worsening of chronic psychiatric conditions [2]. Yet, despite an abundance of anecdotal evidence and peer-reviewed editorials that identify the potential mental health fallout of this public health crisis, the extent of these effects is empirically unknown.

Scientists mobilized quickly to measure many facets of the pandemic, including potential mental health effects. However, the time-consuming and costly nature of survey development [3] and instrument validation make it difficult to draw real-time conclusions [4], especially amid rapidly evolving news cycles that shift pandemic-related discourse. In the absence of survey data, social media represents a potentially valuable data source for studying emergent social issues, including the effect of those issues on behaviors and social mood [5]. Repeated tracking of social media data can provide a diachronic perspective on public morale and collective changes in sentiment, as participants voluntarily contribute to narratives, providing unprompted and diverse understandings of various issues [6-8]. Numerous scholars have successfully used social media data to identify trends and nuances in public mood using a combination of machine learning and artificial intelligence approaches. Some examples include comparing the happiness of users to their online social networks [9,10], identifying detailed predictors of mood through social media feeds [5], predicting cognitive distortions expressed among groups at risk of mental health disorders [11], tracking the emotions of social media users at high resolution [12,13], and mapping negative affectivity among users with internalizing disorders [14]. Collectively, these studies demonstrate the feasibility and value of using sentiment analysis on social media data to study societal mood and well-being, as well as biomedical signals among social media users that can provide useful proxies for mental health [12,15-17]. In fact, these approaches may be especially useful considering the speed that the pandemic became an acute socioeconomic phenomenon, the pervasiveness of COVID-19–related content available online, and the natural reaction of many to post on social media about pandemic-related events.

Indeed, throughout the COVID-19 pandemic, individuals have sought out crisis-related news at increased capacities [18], leading to a collective increase in global social media use [19]. This renders social media data about the COVID-19 pandemic a powerful source of information to draw real-time conclusions about aggregate social well-being during an unprecedented public health event. However, we must remember that, just as survey data are prone to biases, so are data derived from social media [20]. Therefore, to draw accurate inferences about mood, sentiment, and mental health, we must remain cognizant of the type of analysis performed, and what the analysis represents to measure nuanced aspects of sentiment. We contrasted topics discussed, topic-related sentiment, and personal sentiment to arrive at a more comprehensive and accurate assessment of changes in sentiments expressed through social media and their relevance to public health in the United States.

Broadly, this study answers the following three research questions (RQs):

- RQ1: What themes emerge from a corpus of US tweets about COVID-19?
- RQ2: To what extent did social media use increase during the onset of the COVID-19 pandemic?
- RQ3: What patterns emerge from longitudinal tracking of sentiment during the onset of the COVID-19 pandemic?

To address these RQs, we analyzed a large-scale set of Twitter data that are strictly relevant to the topic of COVID-19 in the United States from January 22, 2020, onward. Using this data, we also compiled a second corpus of individual geolocated social media timeline data from the same period to understand changes in personal sentiment as a proxy for mental health and evolving US perceptions of the COVID-19 pandemic.

Methods

Data

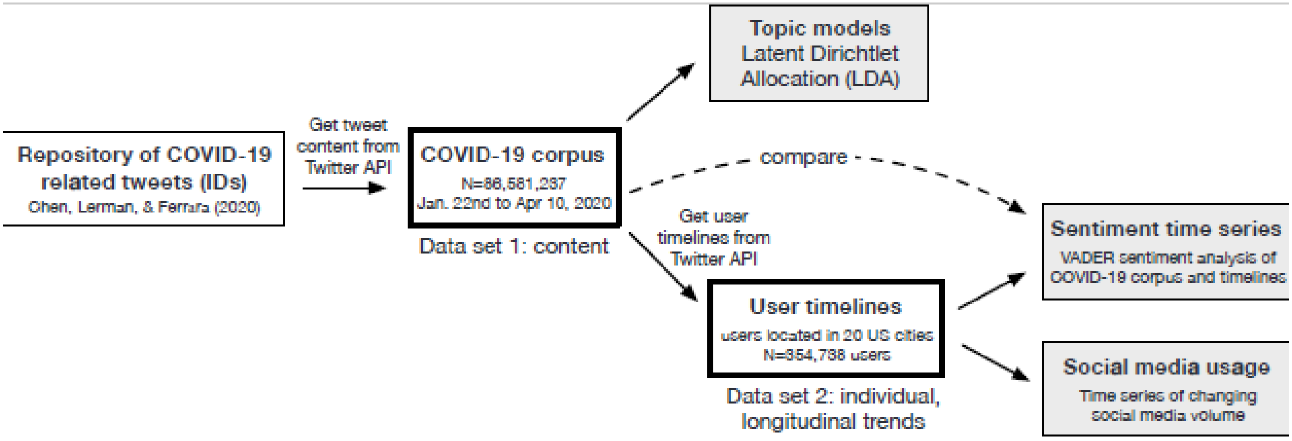
We collected two distinct data sets, each reflecting different aspects of changes in social media behavior before and during the COVID-19 pandemic. The first data set of tweets, collected from an open-access repository containing all COVID-19–related tweets published in the United States [21], was designed to capture *topical differences* (ie, themes) in the Twitter discussion during the events that marked the onset of the pandemic. The repository provides a list of tweet IDs, which we used to extract tweet content from Twitter's application programming interface (API; end point: GET statuses/show/id). We downloaded each tweet as well as the standard metadata provided by Twitter. Specifically, we retrieved COVID-19–related tweets posted between January 22, 2020 (first day of data collection and roughly 1 week prior to the first confirmed US COVID-19 case) through April 9, 2020 (the middle of social distancing efforts). Hereafter, we refer to this set of tweets as the "*COVID-19 corpus*" (n=86,581,237 tweets). Please refer to Figure 1 for a visual representation of this data set and how it was retrieved.

To gauge fluctuations of personal activity and mood at the individual (rather than topical) level, we downloaded the Twitter timelines (ie, the 3200 most recent tweets) of individual social media users who contributed to the COVID-19 corpus and resided within the 20 US cities with the most COVID-19 cases per 100,000 people from the Twitter API (end point: GET statuses/user_timeline). These timelines capture the changing behavior and emotions of individual Twitter users during the

COVID-19 pandemic but do not strictly pertain to tweets exclusively related to COVID-19. We referred to this data as *user timeline data* (n=354,738 users; n=69,349,479 tweets), as shown in Figure 1. All tweets in either data set were scrubbed of any personally identifiable information in accordance with ethical social media use practices.

To ensure that we were measuring expressed sentiment in our data, we excluded non-English tweets and, specifically within the user timeline data, retweets and biasing keywords including "coronavirus," "COVID-19," and "pandemic," among others. These words were removed because they inherently carry a negative connotation, and their inclusion would artificially decrease sentiment given that the corpus itself is composed of COVID-19–related content. In other words, because users are naturally tweeting about *coronavirus*, *virus*, and the *pandemic*, the inclusion of those words may not necessarily reflect the individual's well-being. Note, the resulting sample sizes for each corpora exceeded the mean observed in a recent scoping survey of the literature on social media analytics for public health (n=20,000) [22], resulting in ample representation to conduct our analyses. Additionally, previous studies have used large-scale sentiment analysis to accurately predict social mood [23] and how sentiment expressed on social media correlates with psychological well-being [24]. Thus, the use of sentiment analysis for this study was appropriate.

Figure 1. This diagram illustrates the procedure undertaken to procure Tweet IDs from an open-access COVID-19 repository. These tweet IDs were run through Twitter's API to create two distinct data sets. The COVID-19 corpus, which contains all COVID-19–related English language tweets published in the United States between January-April 2020. The user timeline data contains the 3200 most recent tweets of users residing in the 20 cities most affected by COVID-19. API: application programming interface; VADER: Valence Aware Dictionary for Sentiment Reasoning.



Analyses

Latent Dirichlet Allocation Topic Models

Latent Dirichlet allocation (LDA) topic models are unsupervised machine learning tools that perform probabilistic inferences to consolidate large volumes of text data into manageable themes [25]. Simply, words with high probabilities of association (ie, a high likelihood to appear in proximity with other words) are grouped together to form a latent theme, or topic, that qualitatively represents a content area within the collection of text. These methods have been applied in many ways such as to determine common themes in product reviews [26], to map themes within bodies of scientific literature [27], and to identify themes in social media data [28]. Thus, these tools are appropriate for exploratory analyses that seek to consolidate dense text data.

Sentiment Analysis

Sentiment analysis refers to a set of supervised or unsupervised machine learning and natural language processing techniques that extract affective or emotional indicators from text (eg, to determine whether a tweet expresses a negative or positive emotion about policy [23]). In this study, we used the Valence Aware Dictionary and Sentiment Reasoner (VADER) [29] to gauge the emotional valence of tweets. VADER is a rule-based open-source tool that recognizes common terms, idioms, abbreviations, and jargon while accounting for grammatical structures such as punctuation, negation, hedging, and magnification that are commonly employed in the vernacular of social media platforms. The VADER lexicon is one of the largest of its kind containing over 7500 common terms that are each rated for their emotional valence by 10 independent human raters. However, the word *virus* and its many variations (eg, viruses, viral) are not part of the VADER lexicon, meaning changes in the frequency of these words will not bias VADER scores. VADER has been extensively validated for Twitter content [30], showing some of the highest accuracy and coverage for tweets in a benchmark of more than 20 sentiment analysis tools [31].

Change-Point Detection

We applied the Pruned Exact Linear Time (PELT) change-point detection algorithm to identify significant changes in tweet volume and sentiment [32]. Change-point detection algorithms perform a set of mathematical operations over a time series (a series of time-based observations) to identify points in time where the statistical properties of the time series data changed significantly [33]. The PELT algorithm specifically attempts to find a set of change points for a given time series, such that their number and location in time minimizes a given segmentation cost. We chose the PELT algorithm over other similar change-detection algorithms because it is considered to be a more conservative estimate (preferring not to identify change points unless strict conditions are satisfied), thus yielding more accurate detection of statistical changes [34]. Additionally, PELT uses an offline approach to change detection [34], meaning it can consider all possible data points when identifying significant changes, regardless of the type of data, while maintaining high levels of performance.

Procedure

What Themes Emerge From a Corpus of US Tweets About COVID-19?

We divided the COVID-19 corpus into daily segments and generated one topic model per day consisting of 20 topics each. We chose this number to reflect the widest possible span of themes while summarizing the major themes of the online discussion, which is a process used in previous studies [28,35]. We then looked at the top 20 associated words per topic and collapsed similar words into general themes, taking into account similarities of words (eg, United States and US) and potential misspellings, which are common in social media posts. As an example, *Hubei* and *Wuhan* were collapsed into the theme *China*. We then found the frequency ratio (ie, the number of occurrences of a certain word divided by the total number of words) of COVID-19–related themes (China, United States, pandemic, social distancing, Trump, home, lockdown, and deaths) and plotted them on a daily basis to show the evolution of topics over time, indicating both the contribution of the theme to all content of that day and the relative ranking of these terms among these themes. We used intercoder agreement methods to arrive at mutually agreeable interpretations of collapsed themes [36].

To What Extent Did Social Media Use Increase During the Onset of the COVID-19 Pandemic?

For this analysis we used the user timeline data instead of the general COVID-19 corpus (see Figure 1) because within-subject individual posting frequency is a better marker for tracking changes in social media use behavior [37]. Since Twitter's API limits us to only the 3200 *most recent postings* per individual, we only selected individuals who posted on Twitter before January 22, 2020, retaining 354,738 users. This ensured that our analytic sample captured individual behavior in the 20 most-affected US cities throughout the interval of interest (January 22, 2020, to April 9, 2020). We performed a seasonal decomposition—a method that separates the baseline, trend, and seasonal components of a time series—to determine whether we can observe increased Twitter use during the pandemic relative to events just prior. We then detected significant points of change using the PELT algorithm [38].

What Patterns Emerge From Longitudinal Tracking of Sentiment During the Onset of the COVID-19 Pandemic?

Sentiment can change because individuals discuss different topics (eg, using pejorative terms such as "virus" more frequently) or because of personal, individual changes in how people actually feel. We therefore compared daily VADER sentiment scores for the COVID-19 corpus (to gauge topical sentiment) and the user timeline within-subjects data (to assess personal changes in sentiment) from January 22, 2020 (the first official day of data collection) through April 9, 2020. We determined change points in the time series of daily averaged VADER sentiment with the PELT detection algorithm to identify significant changes in sentiment throughout this time period.

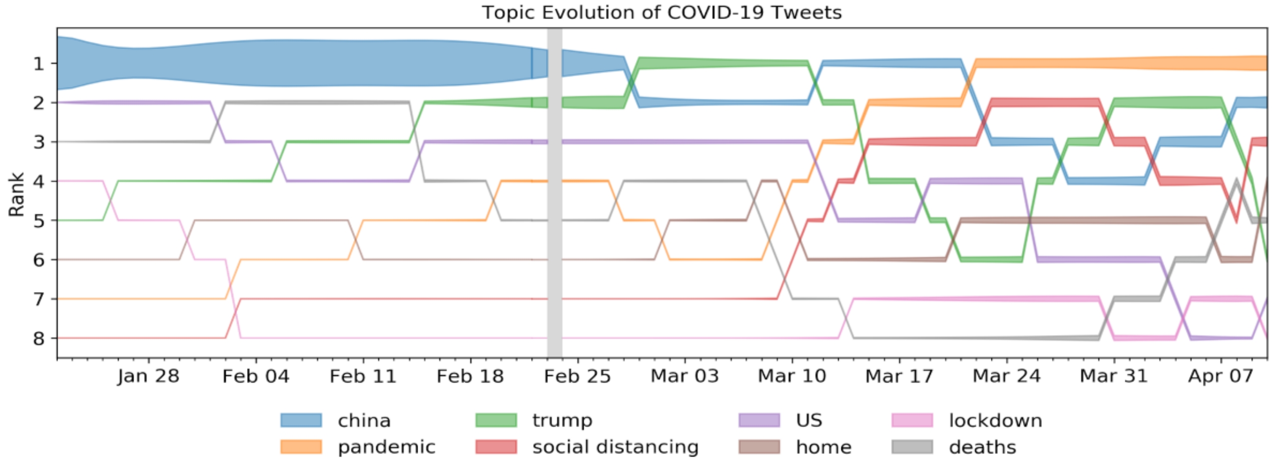
Results

What Themes Emerge From a Corpus of US Tweets About COVID-19?

We consolidated the COVID-19 corpus (n=86,581,237 tweets) into themes using LDA topic models. Figure 1 highlights the eight most salient topics and how their prominence (width of bars and rank) changed over time relative to major COVID-19–related milestones identified by the World Health Organization (WHO).

As shown in Figure 2, topics continued to rise and fall in prominence relative to emerging news cycles throughout the time period under consideration. Indeed, the majority of COVID-19–related Twitter activity focused on China in February 2020. However, from March to April, as the novel coronavirus increasingly began to affect the US population, *China* became less prominent as more US-centered topics such as "lockdown" and "social distancing" emerged. Although China remained a prominent theme throughout the duration of interest, US-centered topics gradually came to dominate social media spaces.

Figure 2. Topic group rankings over time in the COVID-19 corpus. Each topic is ranked by its frequency ratio. The width of the bars indicates the fraction of words in the topic on a given day. The colors of the areas indicate which theme the corresponding area belongs to.

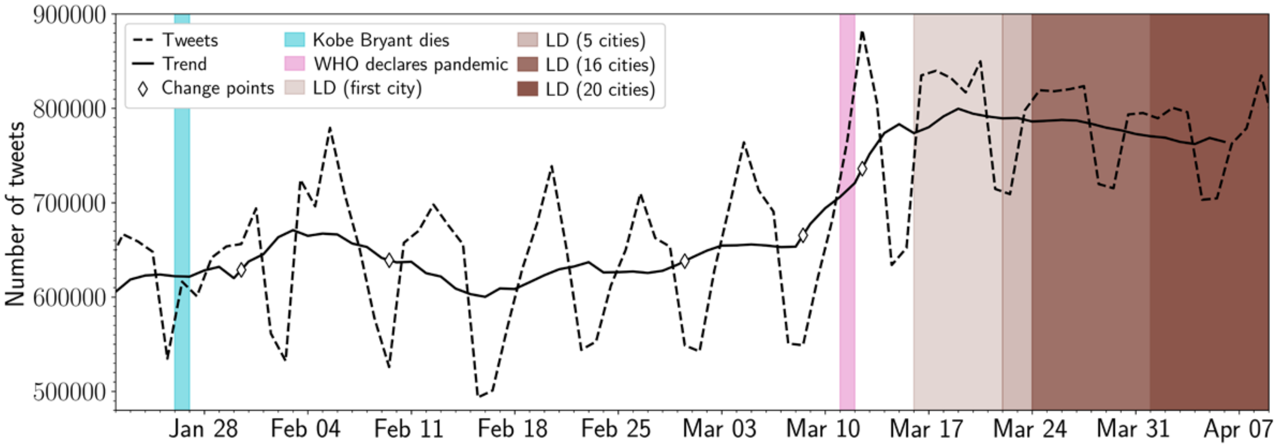


To What Extent Did Social Media Use Increase During the Onset of the COVID-19 Pandemic?

Using user timeline data, we compared the frequency of social media posts before and during the pandemic of 292,000 users (whose timelines span January 22 onwards) in the 20 most-affected metropolitan cities (n= 66,725,505 tweets). Figure 3 highlights changes in posting volume between January 22, 2020, and April 9, 2020. The peaks and troughs (dashed line) of this graph show how seasonal and weekly cycles shape tweet volume. The solid line plots the trend in the time series after removing cycles and seasonal effects (through seasonal decomposition), and the progressively darker shades of brown denote the number of cities that imposed mandatory lockdowns.

Generally, we observed a consistent upward trend in total Twitter volume from early to late March. The PELT change-point algorithm identified two significant volume changes on March 8 and 12, 2020—around the time COVID-19 was declared a global pandemic (March 11) and President Trump declared a national emergency in the United States (March 14). The upward trend stabilized thereafter, albeit at higher observed volumes than prior to the onset of the COVID-19 pandemic. This supports the notion that individuals in our sample were more engaged with social media and made more use of it, possibly to discuss or obtain further information relevant to the news cycle.

Figure 3. Number of daily tweets in a subsample of the user timeline data. The dashed line displays the daily number of tweets in the subsample of the user timeline data. The solid line indicates the underlying trend in the daily number of tweets (after removing its seasonal effects). The diamond markers indicate significant change points indicated by the Pruned Exact Linear Time algorithm. The light blue and pink annotations denote the death of Kobe Bryant (January 26, 2020) and the day when the WHO declared COVID-19 a global pandemic (March 11, 2020), respectively. The brown bars indicate the dates that lockdowns were enforced in the 20 considered cities (ranging from March 16 to April 1, 2020). The opacity of the brown bars indicates how many cities had enforced a lockdown at that date. LD: lockdown; WHO: World Health Organization.



What Patterns Emerge From Longitudinal Tracking of Sentiment During the Onset of the COVID-19 Pandemic?

We applied the VADER sentiment tool to the COVID-19 corpus (to assess sentiment of all US tweets about COVID-19) and the user timeline data (to track changes in user sentiment using their most recent 3200 tweets). Figure 4 tracks sentiment relative to major COVID-19–related milestones for both data sets, with the orange line tracking the COVID-19 corpus and the blue line tracking the user timeline data. In the COVID-19 corpus, there was an unmistakable increase in sentiment with two PELT-identified significant changes on March 9 (just before the WHO classified COVID-19 as a pandemic) and March 19, 2020 (shortly after President Donald Trump declared a national emergency). Figure 5 further shows that the percentage of positively scored COVID-19 tweets increased over time, reinforcing the positive trend.

Conversely, the user timeline data (which again contained the most recent 3200 tweets of a given user) showed decreases in sentiment over the same period. The user timeline data had one PELT-identified significant change in sentiment on January 28, 2020 (the day National Basketball Association [NBA] player Kobe Bryant died in a helicopter crash). There was a notable but short-lived drop in sentiment before March 9 (when the WHO classified COVID-19 as a pandemic).

Figure 4. Distribution of nonzero VADER sentiment in the COVID-19 corpus (orange) and user-timeline data (blue). The solid lines display the 14-day moving average, and orange crosses/blue triangles represent the actual daily average VADER sentiment value. The diamond markers show change points indicated by the Pruned Exact Linear Time algorithm; the color of the diamond's edge refers to the time series that this change point belongs. The light blue, gray, and pink annotations denote the day of Kobe Bryant’s death (January 26, 2020), the day of missing data in the COVID-19 corpus (February 23), and the WHO declaration of COVID-19 as a global pandemic (March 11), respectively. The brown bars indicate the dates that lockdowns were enforced in the 20 considered cities (ranging from March 16 to April 1, 2020). The opacity of the brown bars indicates how many cities had enforced a lockdown at that date. LD: lockdown; VADER: Valence Aware Dictionary for Sentiment Reasoning; WHO: World Health Organization.

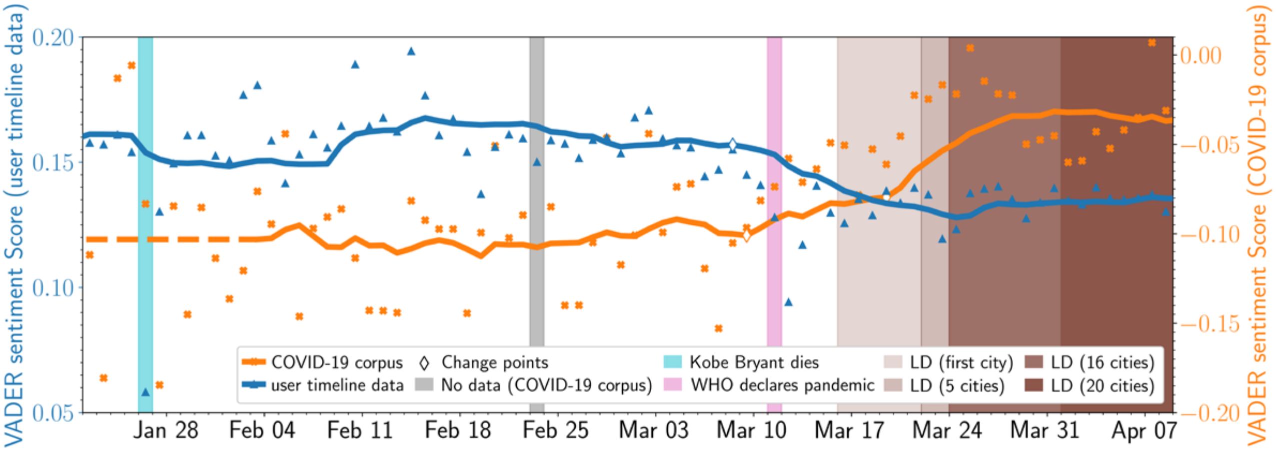
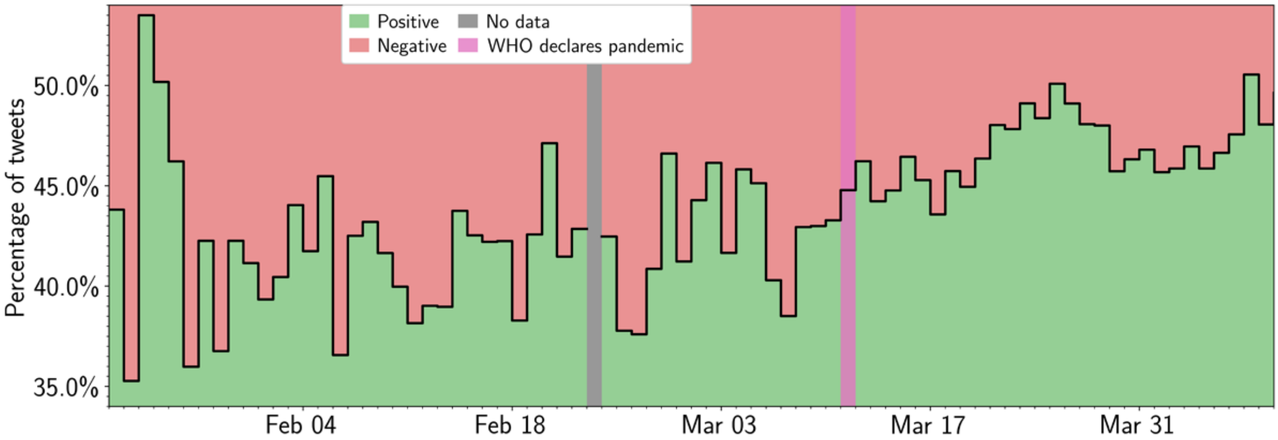


Figure 5. Distribution of positive and negative VADER sentiment in COVID-19 corpus. The solid line displays the fraction of tweets that hold positive or negative VADER sentiment in the user timeline data (colored green for positive and red for negative sentiment). The gray and pink annotations denote the day of missing data (February 23, 2020) and the WHO declaration of COVID-19 as a global pandemic (March 11), respectively. VADER: Valence Aware Dictionary for Sentiment Reasoning; WHO: World Health Organization.



Discussion

Principal Findings

The purpose of this study is to draw conclusions about US mental health amid the COVID-19 pandemic using computational social media analytics of a publicly available repository containing all COVID-19–related tweets. Using the COVID-19 tweet repository released by Chen and colleagues [21], we expanded on their original topical analysis of COVID-19 content by examining both topics and sentiment specific to the US population to understand how the COVID-19 pandemic may be impacting social well-being. We used social media data as the medium for analysis for several reasons, in particular its usefulness to gauge real-time changes in mental health and social well-being before, during, and after rapid socioeconomic changes [15]. In this section, we discuss the findings of this study in detail and highlight public health implications about social well-being during the onset of the COVID-19 pandemic.

COVID-19 Social Media Themes: Content Is Reactionary to News Cycles

Bento and colleagues [18] predicted that crisis-related information seeking would increase during the COVID-19 pandemic. Numerous studies have supported that prediction by identifying increased panic within social media spaces as users react to COVID-19 news–related content on their feeds [39,40]. Through our LDA topic modeling analysis, we echo much of those findings as well; the topics uncovered by LDA frequently corresponded to the life cycles of COVID-19–related news and current events. For example, the name used to identify the virus on social media evolved to match changes in viral terminology as presented on news outlets (eg, the *novel coronavirus* became *coronavirus*, *COVID-19*, and *COVID-19 pandemic*). The locations that emerged within the topics also corresponded to the movement of COVID-19 from mainland China to the United States. For example, in late January, *China* was a dominant theme in Twitter content (shown by the width of the bars in Figure 2), indicating the US public may have considered the then-classified epidemic a foreign matter (eg, tweet: “OMG China just shut down trading...Still think it's just a flu?”). However, in later weeks and months, US-centered topics and tweets replaced *China* in prominence as more Americans became preoccupied with pandemic control measures at home (eg, tweet: “lol it’s wild seeing every single other country grappling with this virus in ways we KNOW the US will never do”). These findings suggest that, as the COVID-19 pandemic became more prominent in the United States, social media content changed to topics of more direct impact to the United States and with greater frequency of use (evidenced by a PELT-identified significant increase in Twitter volume during that time). This volume increase remained consistently above levels observed prior to the COVID-19 pandemic, which is possibly explained by several factors including increased anxiety as the COVID-19 pandemic reached the United States, the isolating effects of statewide stay-at-home orders, or other social fallouts driven by the pandemic. These findings collectively support those of Castillo and colleagues [41], who contend that social media content follows the life cycle of news stories. Thus, news, as a vehicle for crisis-related social media communication, should be studied more intently.

COVID-19 Topical Sentiment: Increased Sentiment May Indicate a Priming Effect

Given the expected negative physical and emotional outcomes of a major pandemic, we were surprised to observe that the sentiment of the COVID-19 corpus trended positively. However, this increase in sentiment is likely not indicative of actual changes in population sentiment but rather the effect of a common *topical* bias in social media analytics [42]. Our COVID-19 data set was specifically selected to contain tweets that related to the topic of COVID-19. Hence, its sentiment scores will reflect the language used to discuss this particular topic and not necessarily the underlying emotion of the population. Topically selected social media posts are likely primed by news cycles [43] and show false optimism—a phenomena where individuals tend to post content that is more positive or optimistic than their true emotions [44]. Previous research has further shown that language in general is biased toward positivity, especially when posts are collected for a random topic [30]. For example, language used to compose a post about COVID-19 may contain trending verbiage or framing devices, including showcasing support of groups frequently part of news cycles (eg, tweet: “Great news...it’s a welcome burden lifted off our incredible nurses and medical first responders”). Similarly, users may just be conveying positivity through carefully selected *popular* words around a trending topic (eg, tweet: “No matter how hard the situation nowadays during the pandemic outbreak, we should keep being positive and optimistic”). Within the VADER lexicon, many of the words commonly used in this context are scored positively (great, welcome, incredible, positive, optimistic), which also artificially inflates sentiment ratings. Thus, we posit that topically driven tweet samples may not validly reflect actual changes in population mood but rather topic-driven language sentiment. This justifies our approach to analyze within-subject timelines of individual posts that are not necessarily bound by the criterion of strictly being COVID-19–related, thereby increasing the odds of reflecting personal changes in mood trajectories.

User Timeline Sentiment: Lower Timeline Sentiment May Indicate Decreased Social Well-being

By contrast, we found a negative trajectory in sentiment scores for the user timeline data. This means that, although content in the COVID-19 corpus trended positively (possibly due to priming), relative to the totality of their timelines, our sample mood was *lower* than it once was. Thus, this comparison gives us deeper insight into *underlying* mood and sentiment. This finding further supports the assertion that the positivity conveyed in social media posts may not validly reflect what people are feeling at a given moment. Indeed, that positivity may be serving as a veneer posted *in the moment* to convey positivity during a time of uncertainty. To obtain an accurate assessment of mood and well-being, other reference points (in this case, social media posts prior to the pandemic) are needed to examine in-the-moment sentiment relative to their prior histories. Within our timeline data, we captured one event that occurred prior to the pandemic that also affected sentiment scores relative to timeline histories—the passing of NBA basketball player Kobe Bryant. The effect of Bryant’s passing led to a sharp decline in sentiment, which lasted approximately 24 hours before returning to levels observed previously. Regarding the pandemic, after the PELT-identified shift on March 8, 2020, sentiment scores (relative to timelines) became not only lower but also consistently lower and did not return to levels observed before the pandemic reached the United States. This trend may hold implications into the longitudinal effects of the pandemic and subsequent impacts on US mood and social sentiment, as this continued trend may be indicative of a long-lasting shift in mood and well-being.

Social Media Use in Times of Crisis

Analyses of aggregated social media feeds are shown to adequately predict other phenomena including the stock market [45]; political leanings [46]; and, when analyzed through a time series, collective shifts in general mood [47]. Our study contributes to this body of literature by highlighting the disparity between how Americans portray themselves on social media versus the latent sentiment they may be experiencing during times of crisis. Generally, Americans were not posting social media content about COVID-19 prior to the first documented US case. Once COVID-19 became a reality in the United States, however, there was a continued increase in the total number of US-based tweets about the novel coronavirus, indicative of growing social media use in our sample. By analyzing the COVID-19 corpus and user timeline data separately, but with the same analytic procedures, we saw divergent findings that reinforced the difference between in-the-moment portrayals versus the longitudinal information that can be gleaned from individual timeline analyses. For example, tweets and posts about COVID-19 may attempt to be lighthearted or convey optimism; however, individually, social media users may not be as optimistic as they were prior to the pandemic. During an unprecedented public health crisis, it is therefore important to look beyond the topical focus of messages on social media that reference the crisis itself as a proxy of public mood, as they are likely to be affected by other influences (eg, political framing and projecting hope). Ultimately, our findings exemplify COVID-19 as a case study in social media behavior, whose outcomes should be generalized to other crisis-related events.

Concluding Remarks and Implications

At the time of writing, the US COVID-19 death toll was just over 136,000, with the lives of millions of people disrupted by the various effects of the pandemic. This study elucidates the possible mental health effects of the COVID-19 pandemic among Twitter users using a computational approach to analyze a corpus of all archived US-based COVID-19 tweets from January to April 2020. These analyses revealed, to varying extents, how the pervasiveness of COVID-19 content available on social media and abrupt shift in lifestyles may be negatively affecting social sentiment relative to points just prior to the pandemic. Given that sentiments expressed on social media have been used as a proxy for mental well-being [48], these findings support calls from public health and medical scholars who contend that mental health is an urgent concern during the COVID-19 pandemic, especially as our findings illustrate a declining trend in sentiment. Thus, we encourage further research on US mental health status amid the pandemic using survey methods or other primary data collections to substantiate our findings with testable outcomes. We also call for more research on mental health interventions amid the COVID-19 pandemic, with particular attention to modality (ie, in person vs virtual), and the efficacy of those efforts.

Limitations

Our study is subject to limitations. Twitter requires users to opt in to geotagging features. Consequently, any information inferred about the user's city of residence is often limited to self-reported data as specified on profile pages. This means that some of our timeline data may not originate from the city specified by a user, as this information can easily be misrepresented [49] (eg, stating they live in New York, NY, but actually residing in Newark, NJ). We also acknowledge a likely bias regarding key demographic information, including age, gender, and socioeconomic status among social media users [50], in addition to the temporal, spatial, and geographic patterns that may affect how sentiment is expressed on social media (eg, older adults posting early in the morning vs younger adults posting late at night or urban vs rural users) [51]. It is also, as of yet, not possible to accurately diagnose someone with a mental health condition through social media feeds alone, although research has shown that social media content contains important indicators with respect to mental health and biomedical signals. Thus, we relied on trend data to draw inferences about the *possibility* of mental health decline based on average sentiment scores [52]. However, these limitations do not diminish the importance or validity of this study. Rather, they create avenues for additional research that expand on the findings of this paper and leverage the limitations inherent to social media data, such as measuring cognitive distortions on Twitter based on posting time or approaches that measure diagnosable mental health conditions through social media data, particularly during times of increased panic and crisis. In addition, because social media has been widely used to draw conclusions about public mood through large-scale sentiment analysis procedures [10,18], we contend our approach is appropriate to draw the conclusions discussed herein. See Multimedia Appendix 1 for source code.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study syntax.
[DOCX File , 28 KB - jmir_v22i12e21418_app1.docx]

References

1. Auerbach J, Miller BF. COVID-19 exposes the cracks in our already fragile mental health system. Am J Public Health 2020 Apr 09:e1-e2. [doi: 10.2105/AJPH.2020.305699] [Medline: 32271609]
2. Pfefferbaum B, North CS. Mental health and the covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
3. Gualano MR, Lo Moro G, Voglino G, Bert F, Siliquini R. Effects of covid-19 lockdown on mental health and sleep disturbances in Italy. Int J Environ Res Public Health 2020 Jul 02;17(13):4779 [FREE Full text] [doi: 10.3390/ijerph17134779] [Medline: 32630821]

4. Coughlan M, Cronin P, Ryan F. Survey research: process and limitations. Int J Ther Rehabil 2009 Jan;16(1):9-15. [doi: 10.12968/ijtr.2009.16.1.37935]
5. Berry N, Emsley R, Lobban F, Bucci S. Social media and its relationship with mood, self-esteem and paranoia in psychosis. Acta Psychiatr Scand 2018 Dec;138(6):558-570 [FREE Full text] [doi: 10.1111/acps.12953] [Medline: 30203454]
6. Cho SE, Jung K, Park HW. Social media use during Japan's 2011 earthquake: how Twitter transforms the locus of crisis communication. Media Int Aust 2013 Nov;149(1):28-40. [doi: 10.1177/1329878x1314900105]
7. Page RE. Stories and Social Media: Identities and Interaction. Abingdon, UK: Routledge; 2013.
8. Stieglitz S, Dang-Xuan L. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. J Manage Inf Syst 2014 Dec 08;29(4):217-248. [doi: 10.2753/MIS0742-1222290408]
9. Bollen J, Gonçalves B, Ruan G, Mao H. Happiness is assortative in online social networks. Artif Life 2011;17(3):237-251. [doi: 10.1162/artl_a_00034] [Medline: 21554117]
10. Bollen J, Gonçalves B, van de Leemput I, Ruan G. The happiness paradox: your friends are happier than you. EPJ Data Sci 2017 May 18;6(1):1-10. [doi: 10.1140/epjds/s13688-017-0100-1]
11. Simms T, Ramstedt C, Rich M, Richards M, Martinez T, Giraud-Carrier C. Detecting cognitive distortions through machine learning text analytics. 2017 Presented at: 2017 IEEE International Conference on Healthcare Informatics (ICHI); August 2017; Park City, UT. [doi: 10.1109/ichi.2017.39]
12. Correia RB, Wood IB, Bollen J, Rocha LM. Mining social media data for biomedical signals and health-related behavior. Annu Rev Biomed Data Sci 2020 Jul;3:433-458 [FREE Full text] [doi: 10.1146/annurev-biodatasci-030320-040844] [Medline: 32550337]
13. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881 [FREE Full text] [doi: 10.1126/science.1202775] [Medline: 21960633]
14. Johns Hopkins University Bloomberg School of Public Health. Social media use by adolescents linked to internalizing behaviors. Medical Xpress. 2019 Sep 11. URL: https://medicalxpress.com/news/2019-09-social-media-adolescents-linked-internalizing.html [accessed 2020-07-16]
15. Bollen J, Pepe A, Mao H. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. arXiv. 2009 Nov 09. URL: http://arxiv.org/abs/0911.1583 [accessed 2020-05-31]
16. Dodds PS, Clark EM, Desu S, Frank MR, Reagan AJ, Williams JR, et al. Reply to Garcia et al.: common mistakes in measuring frequency-dependent word characteristics. Proc Natl Acad Sci U S A 2015 Jun 09;112(23):E2984-E2985 [FREE Full text] [doi: 10.1073/pnas.1505647112] [Medline: 25997446]
17. Bekalu MA, McCloud RF, Viswanath K. Association of social media use with social well-being, positive mental health, and self-rated health: disentangling routine use from emotional connection to use. Health Educ Behav 2019 Dec;46(2_suppl):69-80. [doi: 10.1177/1090198119863768] [Medline: 31742462]
18. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn YY, Simon K. Information seeking responses to news of local COVID-19 cases: evidence from internet search data. arXiv. Preprint posted online April 6, 2020 [FREE Full text] [doi: 10.1073/pnas.2005335117]
19. Koeze E, Popper N. The virus changed the way we internet. The New York Times. 2020 Apr 07. URL: https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html [accessed 2020-05-31]
20. Olteanu A, Castillo C, Diaz F, Kıcıman E. Social data: biases, methodological pitfalls, and ethical boundaries. Front Big Data 2019 Jul 11;2:13. [doi: 10.3389/fdata.2019.00013]
21. Chen E, Lerman K, Ferrara E. COVID-19: the first public coronavirus Twitter dataset. Preprint posted online March 16, 2020 [FREE Full text]
22. Edo-Osagie O, De La Iglesia B, Lake I, Edeghere O. A scoping review of the use of Twitter for public health research. Comput Biol Med 2020 Jul;122:103770 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103770] [Medline: 32502758]
23. Liu B. Sentiment analysis and opinion mining. Synthesis Lect Hum Lang Technologies 2012 May 23;5(1):1-167. [doi: 10.2200/S00416ED1V01Y201204HLT016]
24. Jaidka K, Giorgi S, Schwartz HA, Kern ML, Ungar LH, Eichstaedt JC. Estimating geographic subjective well-being from Twitter: a comparison of dictionary and data-driven language methods. Proc Natl Acad Sci U S A 2020 May 12;117(19):10165-10171 [FREE Full text] [doi: 10.1073/pnas.1906364117] [Medline: 32341156]
25. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Machine Learning Res 2003 Jan;3:993-1022.
26. Wang H, Ding Y, Tang J, Dong X, He B, Qiu J, et al. Finding complex biological relationships in recent PubMed articles using Bio-LDA. PLoS One 2011 Mar 23;6(3):e17243 [FREE Full text] [doi: 10.1371/journal.pone.0017243] [Medline: 21448266]
27. Moghaddam S, Ester M. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In: Proceedings of the 34th international ACM SIGIR Conference on Research and Development in Information Retrieval. 2011 Presented at: SIGIR '11; July 2011; Beijing, China p. 665-674. [doi: 10.1145/2009916.2010006]
28. Barry AE, Valdez D, Padon AA, Russell AM. Alcohol advertising on Twitter—a topic model. Am J Health Education 2018 Jun 29;49(4):256-263. [doi: 10.1080/19325037.2018.1473180]
29. Hutto CJ, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 2014; Ann Arbor, MI.

30. Elbagir S, Yang J. Twitter sentiment analysis using natural language toolkit and VADER sentiment. In: Proceedings of the International MultiConference of Engineers and Computer Scientists. 2019 Presented at: IMECS 2019; March 2019; Hong Kong.
31. Ribeiro MT, Singh S, Guestrin C. Model-agnostic interpretability of machine learning. arXiv. 2016 Jun 16. URL: http://arxiv.org/abs/1606.05386 [accessed 2020-05-19]
32. Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. J Am Stat Assoc 2012 Oct 17;107(500):1590-1598. [doi: 10.1080/01621459.2012.737745]
33. Liu S, Yamada M, Collier N, Sugiyama M. Change-point detection in time-series data by relative density-ratio estimation. Neural Netw 2013 Jul;43:72-83. [doi: 10.1016/j.neunet.2013.01.012] [Medline: 23500502]
34. Wambui GD, Waititu GA, Wanjoya A. The power of the Pruned Exact Linear Time(PELT) test in multiple changepoint detection. Am J Theor Appl Stat 2015;4(6):581. [doi: 10.11648/j.ajtas.20150406.30]
35. Valdez D, Pickett AC, Goodson P. Topic modeling: latent semantic analysis for the social sciences. Soc Sci Q 2018 Sep 07;99(5):1665-1679. [doi: 10.1111/ssqu.12528]
36. Kuckartz U, Rädiker S. Analyzing intercoder agreement. In: Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video. New York, NY: Springer International Publishing; 2019:267-282.
37. Lee K, Agrawal A, Choudhary A. Real-time disease surveillance using Twitter data: demonstration on flu and cancer. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2013 Presented at: KDD '13; August 2013; Chicago, Il p. 1474-1477. [doi: 10.1145/2487575.2487709]
38. Gabrielli L, Rinzivillo S, Ronzano F, Villatoro D. From tweets to semantic trajectories: mining anomalous urban mobility patterns. In: Nin J, Villatoro D, editors. Citizen in Sensor Networks. New York, NY: Springer International Publishing; 2014:26-35.
39. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
40. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):taaa031 [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
41. Castillo C, El-Haddad M, Pfeffer J, Stempeck M. Characterizing the life cycle of online news stories using social media reactions. In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing. 2014 Presented at: CSCW '14; February 2014; Baltimore, MD p. 211-223. [doi: 10.1145/2531602.2531623]
42. Ferrara E, Yang Z. Measuring emotional contagion in social media. PLoS One 2015;10(11):e0142390 [FREE Full text] [doi: 10.1371/journal.pone.0142390] [Medline: 26544688]
43. Schleuder JD, White AV, Cameron GT. Priming effects of television news bumpers and teasers on attention and memory. J Broadcasting Electronic Media 1993;37:437-452.
44. Bala K. Social media and changing communication patterns. Global Media J-Indian Edition 2014 Jun;5(1):1-6.
45. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. J Computational Sci 2011 Mar;2(1):1-8. [doi: 10.1016/j.jocs.2010.12.007]
46. Makazhanov A, Rafiei D, Waqar M. Predicting political preference of Twitter users. Soc Network Analysis Mining 2014 May 6;4(1):193. [doi: 10.1007/s13278-014-0193-5]
47. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: analyzing the language of mental health on Twitter through self-reported diagnoses. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 2015; Denver, CO. [doi: 10.3115/v1/w15-1201]
48. Derks D, Fischer AH, Bos AE. The role of emotion in computer-mediated communication: a review. Comput Hum Behav 2008 May;24(3):766-785. [doi: 10.1016/j.chb.2007.04.004]
49. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
50. Smith A, Brenner J. Twitter Use 2012. Pew Research Center. 2012 May 31. URL: https://www.pewresearch.org/internet/2012/05/31/twitter-use-2012/ [accessed 2020-09-17]
51. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
52. De Choudhury M, De S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 2014; Ann Arbor, MI.

Abbreviations

API: application programming interface
LDA: latent Dirichlet allocation
NBA: National Basketball Association

PELT: Pruned Exact Linear Time
RQ: research question
VADER: Valence Aware Dictionary and Sentiment Reasoner
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.06.20; peer-reviewed by Z Zhang, R Gore, M Alvarez de Mon; comments to author 13.07.20; revised version received 20.07.20; accepted 07.12.20; published 14.12.20.

Please cite as:
Valdez D, ten Thij M, Bathina K, Rutter LA, Bollen J
Social Media Insights Into US Mental Health During the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data
J Med Internet Res 2020;22(12):e21418
URL: http://www.jmir.org/2020/12/e21418/
doi:10.2196/21418
PMID:33284783

Viewpoint

Where Virtual Care Was Already a Reality: Experiences of a Nationwide Telehealth Service Provider During the COVID-19 Pandemic

Lori Uscher-Pines[1], MSc, PhD; James Thompson[2], BA; Prentiss Taylor[2], MD; Kristin Dean[2], MD; Tony Yuan[2], MD; Ian Tong[2], MD; Ateev Mehrotra[3], MD, MPH

[1]RAND Corporation, Health Care Division, Arlington, VA, United States
[2]Doctor on Demand, San Francisco, CA, United States
[3]Harvard Medical School, Boston, MA, United States

Corresponding Author:
Lori Uscher-Pines, MSc, PhD
RAND Corporation
Health Care Division
1200 S Hayes St
Arlington, VA, 22202
United States
Phone: 1 7034131100
Email: luscherp@rand.org

Abstract

Background: The COVID-19 pandemic has led to an increase in the use of and demand for telehealth services.

Objective: Here, we describe the utilization of telehealth services provided by Doctor On Demand, Inc., a well-known telehealth company in the United States, before and during the COVID-19 pandemic. We also explore how the number of virtual visits, reasons for visits, and patients served changed over time.

Methods: We reported data as a percentage change from the baseline week during 2 distinct time periods: February-June 2019 and February-June 2020 based on 4 categories of visits: respiratory illness, unscheduled behavioral health, scheduled behavioral health, and chronic illness.

Results: In 2020, the total visit volume increased considerably from March through April 7, 2020 (59% above the baseline) and then declined through the week of June 2 (15% above the baseline). Visits for respiratory illnesses increased through the week of March 24 (30% above the baseline) and then steadily declined through the week of June 2 (65% below the baseline). Higher relative increases were observed for unscheduled behavioral health and chronic illness visits through April (109% and 131% above the baseline, respectively) before a decline through the week of June 2 (69% and 37% above the baseline, respectively). Increases in visit volume among rural residents were slightly higher than those among urban residents (peak at 64% vs 58% above the baseline, respectively).

Conclusions: Although this telehealth service provider observed a substantial increase in the volume of visits during the COVID-19 pandemic, it is interesting to note that this growth was not fueled by COVID-19 concerns but by visits for behavioral health and chronic illness. Telehealth services may play a role as a “safety valve” for patients who have difficulty accessing care during a public health emergency.

(J Med Internet Res 2020;22(12):e22727) doi:10.2196/22727

KEYWORDS

telehealth; telemedicine; COVID-19; pandemic; infectious disease; virus; United States

Introduction

In response to the COVID-19 pandemic, health care delivery in the United States has changed dramatically since March 2020. As a result of stay-at-home orders, many physicians who had limited prior experience with telehealth started offering telehealth visits to support social distancing, conserve the use of personal protective equipment, and safeguard vulnerable patients from exposure to COVID-19. This transformation in health care delivery was unprecedented, and evidence about clinicians' experiences transitioning to and maintaining telehealth services during the pandemic is still emerging.

However, when COVID-19 outbreaks emerged, many organizations were already conducting telehealth visits in high volumes. Prior to 2020, large telehealth service providers were providing millions of telehealth visits per year, offering patients immediate access to clinicians via videoconferencing visits from personal electronic devices [1].

To date, research on telehealth implementation during the COVID-19 pandemic has focused on the experiences of professionals new to telehealth [2]. Little is known about the impact of COVID-19 among telehealth services that were functional prior to the pandemic. To address this gap in the literature, we collaborated with a prominent telehealth service provider, Doctor On Demand, Inc., to describe their experience during the ongoing pandemic and to explore how the number of telehealth visits, reasons for visits, and patients served have changed over time.

Methods

Services Offered

Doctor On Demand is a telehealth company that provides services across all 50 US states. It delivers urgent care, behavioral health, preventive care, and chronic care services directly to consumers through its affiliations with self-insured employers and health plans. In March 2020, the company observed an increase in requests for visits and pursued several strategies to increase its capacity. They launched an online COVID-19 assessment tool and information center for mobile devices. The assessment tool, which was developed in collaboration with the Centers for Disease Control and Prevention, allowed patients to compare their symptoms to those suggestive of COVID-19 and provided recommendations regarding whether they need to seek care. The company also hired additional providers and increased the working hours of existing providers. Visits during the study period were delivered by board-certified primary care physicians, board-certified psychiatrists, and doctoral-level psychologists.

Data Analysis

The company generated data in an aggregate form, as a percentage change from the baseline over 2 distinct time periods: February-June 2019 and February-June 2020. The baseline week was defined as February 25 to March 3 for 2019 and February 24 to March 1 for 2020. We selected this specific week-long period because it represented the tail end of the influenza season and, in 2020, it occurred before significant community transmission of COVID-19 was reported in the United States.

We plotted weekly changes in the visit volume from the baseline for all virtual visits and 4 specific categories of visits: respiratory illness (including acute respiratory infection, influenza-like illness, and potential COVID-19), unscheduled behavioral health services offered within the urgent care service staffed by primary care providers, scheduled behavioral health services (including therapy and psychiatry) offered within the behavioral health service staffed by specialty behavioral health providers, and chronic illness.

In the respiratory illness category, we combined acute respiratory infection, influenza-like illness, and suspected COVID-19 given the similarities in their presentation with fever and potential cough. Respiratory illness visits included visits for the following diagnoses: coronavirus infection; viral infection, unspecified; SARS-associated coronavirus; other coronavirus; acute nasopharyngitis; acute maxillary sinusitis; acute frontal sinusitis; acute pansinusitis; acute sinusitis; acute recurrent sinusitis; streptococcal pharyngitis; acute pharyngitis; acute tonsillitis; acute laryngitis; acute obstructive laryngitis (croup); acute upper respiratory infection, unspecified; influenza; viral pneumonia, unspecified; pneumonia; acute bronchitis; acute bronchiolitis; acute lower respiratory infection, unspecified; allergic rhinitis; chronic rhinitis; bronchitis; chronic bronchitis, unspecified; asthma, unspecified; acute bronchospasm; cough; dyspnea; shortness of breath; wheezing; nasal congestion; other disturbances of smell or taste; fever, unspecified; headache; other fatigue; and COVID-19 acute respiratory disease.

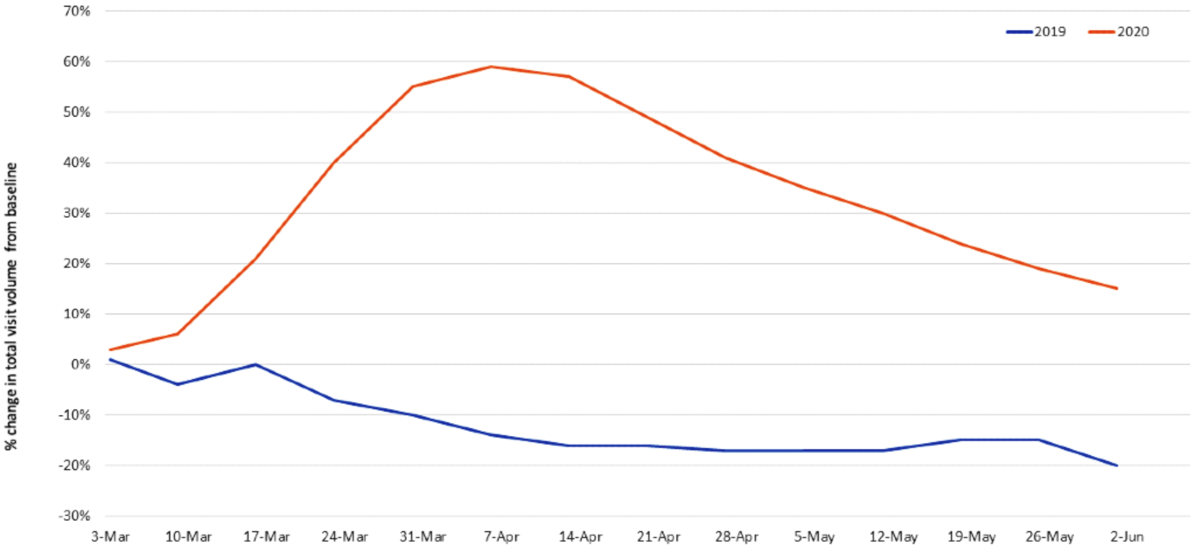
Chronic illness visits included visits for the following diagnoses: asthma, low back pain, hypertension, thyroid disorders, obesity, hypercholesterolemia/hyperlipidemia, chronic obstructive pulmonary disease, type 2 diabetes, complex diabetes, osteoarthritis, iron deficiency anemia, rheumatoid arthritis, fibromyalgia, prediabetes, lupus, ulcerative colitis, heart disease, Crohn disease, HIV, cancer, emphysema, sleep apnea, glaucoma, hemophilia, falls, hepatitis C, chronic kidney disease, end-stage liver disease, scleroderma, stroke, severe chronic kidney disease, hepatitis B, hemochromatosis, macrocytic anemia, macrocytosis, polymyalgia rheumatica, hepatitis A, joint replacement, and interstitial lung disease.

Results

Visit Volume

Compared to the baseline week, in 2019, total visit volume declined from March through June 2019 (Figure 1). In contrast, in 2020, total visit volume increased sharply from March to April 7, 2020 (59% above the baseline) and then steadily declined through the week of June 2, 2020 (15% above the baseline).

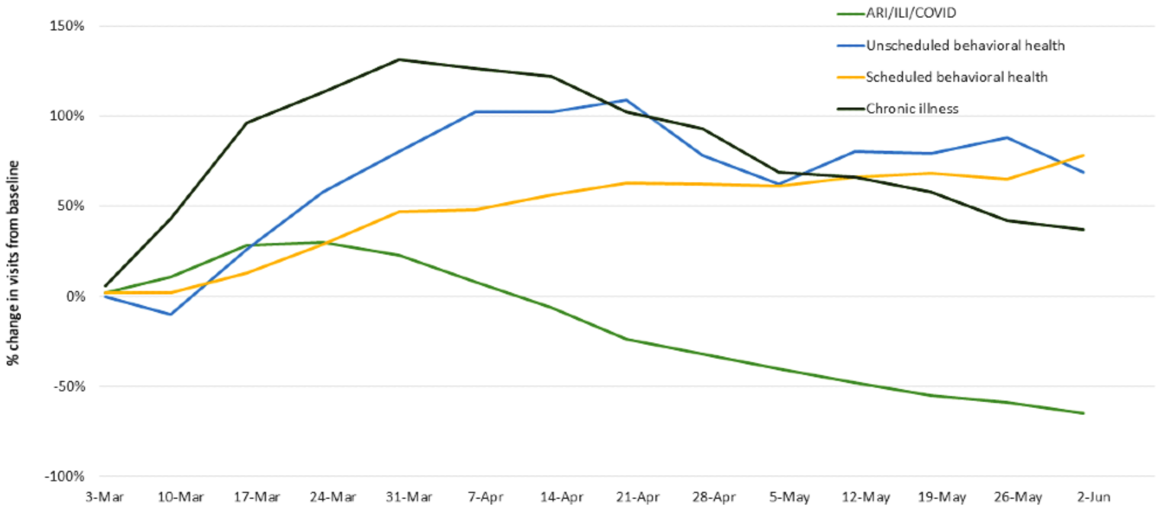
Figure 1. Percentage change in the total volume of virtual visits from the baseline week in 2019 and 2020. Baseline weeks: February 25 to March 3, 2019, and February 24 to March 1, 2020.



In 2020, during the baseline week starting February 24, respiratory illness visits represented 45% of the total visit volume, whereas the total number of visits for behavioral health (scheduled and unscheduled) and chronic illness comprised 20% and 5% of the total visit volume, respectively.

Moreover, in 2020, visits for respiratory illnesses modestly increased initially (peak at 30% above the baseline in the week of March 24) and then steadily declined (65% below the baseline in the week of June 2; Figure 2). In contrast, unscheduled behavioral health and chronic illness visits increased across this period, peaking at 109% in the week of April 21 and at 131% in the week of March 31, before declining to 69% and 37%, respectively—levels that were still above the baseline. In the week of June 2, 2020, respiratory illness visits represented 14% of the total visit volume, whereas behavioral health (scheduled and unscheduled) and chronic illness visits comprised 31% and 5%, respectively.

Figure 2. Percentage change in the total volume of virtual visits by type of visit from the baseline week in 2020. ARI: acute respiratory infection, ILI: influenza-like illness.



Patients Served

In 2020, differences in visit trends by patient location were observed. All visits among urban residents peaked at 58% above baseline, whereas visits among rural residents peaked at 64% above baseline.

Individuals residing in low-income regions (mean per capita income of <US $20,000) accounted for 47% of all visits in January and February 2020 and 50% of all visits in April 2020. The proportion of patients new to the telehealth platform increased from 40% in February 2020 to 53% in April 2020.

Discussion

Consistent with health care providers who were not focused on telehealth prior to the COVID-19 pandemic [2], the telehealth provider Doctor On Demand experienced a substantial increase in the total visit volume during the ongoing pandemic. It is noteworthy that this growth is not attributed to COVID-19 concerns; instead, behavioral health and chronic illness visits seemed to have contributed to the growth.

Although the overall growth of 59% above the baseline is substantial and similar to that reported by other telehealth services [3], it is significantly lower than the telehealth growth rate reported among in-person providers during the pandemic. For In-person providers, telehealth grew from <1% of visits to 14%-43% of visits, which corresponds to growth of >1000% [2,4]. The differences in relative growth likely suggest that the use of telehealth services among in-person providers was very low at baseline.

The overall number of telehealth visits on the Doctor On Demand platform peaked at approximately the same time as emergency department visits in the USA were at their lowest point [5]. This finding suggests that the demand for telehealth in April 2020 may have been driven in part by patient hesitation to seek in-person care. Increasing comfort with in-person care may result in a reduced demand for telehealth services as the pandemic progresses. Studies have also found that the use of telehealth services has reduced as in-person visits have rebounded [6]. However, the exposure to telehealth in the spring of 2020 may result in an increased use of these services over the long-term.

It is unclear whether the increased demand for behavioral health visits during the study period was driven by a higher incidence of mental health concerns due to pandemic-related stressors (eg, increased isolation and financial hardship) or was attributed to the reduced capacity of health care providers practicing in the community. Telehealth services may serve as a “safety valve” for patients, addressing gaps in access to the traditional (in-person) health care delivery system. Previous studies have highlighted the important role that telehealth plays during a pandemic, helping protect patients and clinicians from exposure to disease and maintaining continuity of care [7,8]. National telehealth services are especially well positioned to respond to local or regional emergencies given they have large panels of providers who are spread out geographically, and as such they can support load balancing across their networks [7]. However, in a nation affected by a pandemic, the capacity of these telehealth services is somewhat constrained by the fact that they need to hire and train additional providers to meet unexpected surges in the demand for health care. To increase the preparedness of telehealth services to scale up quickly in a national emergency, it may be necessary to maintain “reserve” providers who can be activated on a short notice.

There has been substantial discussion about the “digital divide” in health care and concerns that patients from lower-income communities may not have the necessary technology or digital literacy to participate in video-based visits. Nevertheless, it is reassuring that the proportion of telehealth visits by patients from lower-income communities was stable during the early weeks of the COVID-19 pandemic.

Conclusions

In summary, we observed that the substantial growth in the use of telehealth services of Doctor On Demand, during the pandemic was surprisingly driven by the increase in visits for behavioral health and chronic illness. Future studies should continue to describe changes in the volume of telehealth visits and reasons for their use as the current COVID-19 pandemic progresses.

Conflicts of Interest

JT, PT, KD, TY, and IT are all employees of Doctor on Demand. PT is also a small shareholder in the company, and IT is the Chief Medical Officer of Doctor on Demand and also a stock holder. All other authors declare no conflict of interest.

References

1. Teladoc Health, Inc. Annual Report Pursuant to Section 13 or 15(D) of the Securities Exchange Act of 1934. United States Securities and Exchange Commission. 2019. URL: http://d18rn0p25nwr6d.cloudfront.net/CIK-0001477449/733d2845-dc97-44f3-bff2-8e7e1c863e81.pdf [accessed 2020-11-10]
2. Mehrotra A, Chernew M, Linetsky D, Hatch H, Cutler D. The Impact of the COVID-19 Pandemic on Outpatient Visits: A Rebound Emerges. The Commonwealth Fund. 2020 May 19. URL: https://www.commonwealthfund.org/publications/2020/apr/impact-covid-19-outpatient-visits [accessed 2020-11-10]
3. Krouse S. Teladoc's Remote Doctor Visits Surge in Coronavirus Crisis. The Wall Street Journal. 2020 Apr 14. URL: https://www.wsj.com/articles/teladocs-remote-doctor-visits-surge-in-coronavirus-crisis-11586894400 [accessed 2020-11-10]
4. Bosworth A, Ruhter J, Samson LW, Sheingold S, Taplin C, Tarazi W, et al. Medicare Beneficiary Use of Telehealth Visits: Early Data from the Start of the COVID-19 Pandemic. Office of the Assistant Secretary for Planning and Evaluation, U.S.

Department of Health and Human Services. Washington, DC: U.S. Department of Health and Human Services; 2020 Jul 28. URL: https://aspe.hhs.gov/system/files/pdf/263866/HP_IssueBrief_MedicareTelehealth_final7.29.20.pdf [accessed 2020-11-10]

5. Hartnett KP, Kite-Powell A, DeVies J, Coletta MA, Boehmer TK, Adjemian J, National Syndromic Surveillance Program Community of Practice. Impact of the COVID-19 Pandemic on Emergency Department Visits - United States, January 1, 2019-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 12;69(23):699-704 [FREE Full text] [doi: 10.15585/mmwr.mm6923e1] [Medline: 32525856]
6. Mehrotra A, Linetsky D, Hatch H. This is supposed to be telemedicine's time to shine. Why are doctors abandoning it? STAT. 2020 Jun 25. URL: https://www.statnews.com/2020/06/25/telemedicine-time-to-shine-doctors-abandoning-it/ [accessed 2020-11-10]
7. Uscher-Pines L, Fischer S, Tong I, Mehrotra A, Malsberger R, Ray K. Virtual First Responders: the Role of Direct-to-Consumer Telemedicine in Caring for People Impacted by Natural Disasters. J Gen Intern Med 2018 Aug;33(8):1242-1244 [FREE Full text] [doi: 10.1007/s11606-018-4440-8] [Medline: 29691713]
8. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]

Edited by G Eysenbach; submitted 21.07.20; peer-reviewed by K Martinez, I Kedan; comments to author 11.08.20; revised version received 20.08.20; accepted 25.10.20; published 15.12.20.

Please cite as:
Uscher-Pines L, Thompson J, Taylor P, Dean K, Yuan T, Tong I, Mehrotra A
Where Virtual Care Was Already a Reality: Experiences of a Nationwide Telehealth Service Provider During the COVID-19 Pandemic
J Med Internet Res 2020;22(12):e22727
URL: https://www.jmir.org/2020/12/e22727
doi:10.2196/22727
PMID:33112761

Original Paper

Computing SARS-CoV-2 Infection Risk From Symptoms, Imaging, and Test Data: Diagnostic Model Development

Christopher D'Ambrosia[1], MS; Henrik Christensen[1], PhD; Eliah Aronoff-Spencer[2], MD, PhD

[1]Department of Computer Science and Engineering, University of California San Diego, San Diego, CA, United States

[2]Division of Infectious Diseases and Global Public Health, School of Medicine, University of California San Diego, San Diego, CA, United States

Corresponding Author:
Eliah Aronoff-Spencer, MD, PhD
Division of Infectious Diseases and Global Public Health
School of Medicine
University of California San Diego
9500 Gilman Drive 0711
San Diego, CA, 92101
United States
Phone: 1 6462348153
Email: earonoffspencer@health.ucsd.edu

Abstract

Background: Assigning meaningful probabilities of SARS-CoV-2 infection risk presents a diagnostic challenge across the continuum of care.

Objective: The aim of this study was to develop and clinically validate an adaptable, personalized diagnostic model to assist clinicians in ruling in and ruling out COVID-19 in potential patients. We compared the diagnostic performance of probabilistic, graphical, and machine learning models against a previously published benchmark model.

Methods: We integrated patient symptoms and test data using machine learning and Bayesian inference to quantify individual patient risk of SARS-CoV-2 infection. We trained models with 100,000 simulated patient profiles based on 13 symptoms and estimated local prevalence, imaging, and molecular diagnostic performance from published reports. We tested these models with consecutive patients who presented with a COVID-19–compatible illness at the University of California San Diego Medical Center over the course of 14 days starting in March 2020.

Results: We included 55 consecutive patients with fever (n=43, 78%) or cough (n=42, 77%) presenting for ambulatory (n=11, 20%) or hospital care (n=44, 80%). In total, 51% (n=28) were female and 49% (n=27) were aged <60 years. Common comorbidities included diabetes (n=12, 22%), hypertension (n=15, 27%), cancer (n=9, 16%), and cardiovascular disease (n=7, 13%). Of these, 69% (n=38) were confirmed via reverse transcription-polymerase chain reaction (RT-PCR) to be positive for SARS-CoV-2 infection, and 20% (n=11) had repeated negative nucleic acid testing and an alternate diagnosis. Bayesian inference network, distance metric learning, and ensemble models discriminated between patients with SARS-CoV-2 infection and alternate diagnoses with sensitivities of 81.6%-84.2%, specificities of 58.8%-70.6%, and accuracies of 61.4%-71.8%. After integrating imaging and laboratory test statistics with the predictions of the Bayesian inference network, changes in diagnostic uncertainty at each step in the simulated clinical evaluation process were highly sensitive to location, symptom, and diagnostic test choices.

Conclusions: Decision support models that incorporate symptoms and available test results can help providers diagnose SARS-CoV-2 infection in real-world settings.

(J Med Internet Res 2020;22(12):e24478) doi:10.2196/24478

KEYWORDS

health; informatics; computation; COVID-19; infection; risk; symptom; imaging; diagnostic; probability; machine learning; Bayesian; model

Introduction

Despite advances in molecular diagnostics and imaging, ruling in or ruling out COVID-19 infection in an individual patient remains a significant challenge [1]. Current guidelines from the Centers for Disease Control and Prevention (CDC) recommend providers to determine whether signs or symptoms are compatible with COVID-19 infection and to test appropriate patients using nucleic acid amplification tests (NAATs) or antigen detection assays [2]. However, the diverse clinical presentations of COVID-19 infection may mimic those of common infections, potentially confounding the diagnostic value of presenting symptoms [3]. Moreover, significant and evolving differences in estimated local disease prevalence for both COVID-19 infection and seasonal respiratory illnesses meaningfully impact differential diagnostic probabilities. Despite the uncertain utility of this symptom and local prevalence information, in low-resource and community settings such as ambulatory clinics, nursing homes, and telemedicine, these may be the only sources of data. In higher-resource settings, NAATs [4], antibody based lateral flow assays [5], chest radiography (CXR) [6], and computed tomography (CT) [7] may be available, yet published literature notes varied performance. Despite these limitations, clinicians with access to any imaging and testing modalities must optimize diagnostic imaging and testing sequences to appropriately reduce diagnostic uncertainty in a given setting.

This complexity highlights the need for reliable and user-friendly clinical decision support systems (CDSS) that suggest optimal testing strategies and quantify SARS-CoV-2 infection risk for patients across the continuum of care. Prior research has demonstrated the potential utility of Bayesian inference [8,9] and machine learning [10,11] methods in diagnostic decision making, but computational clinical decision support has often been underutilized due to a lack of accessibility, transparency, workflow integration, and most importantly, the flexibility to incorporate local provider beliefs into the diagnostic model [12,13].

A robust diagnostic risk model should be built on individualized patient data that is easily obtained by patients and health care workers. Menni et al [14] analyzed a large database of smartphone-enabled, self-reported symptom tracker records to predict potential COVID-19 cases using logistic regression models. In the US test set, this approach had a reported sensitivity of 66% and a specificity of 83%. Ahsan et al [15] used deep learning techniques to differentiate between COVID-19 and non–COVID-19 patients based on open-source imaging and clinical data. However, the performance of this or other machine learning models in clinical settings has not yet been examined.

Moreover, in evolving contexts where illness presentation may change depending on host and viral characteristics, large databases of individual patient records may not be available or locally relevant. Constructing inflexible predictive algorithms, such as logistic regression models, based on out-of-date and locally irrelevant data sets would significantly compromise diagnostic accuracy. Addressing these issues, Chishti et al [16] demonstrated the advantages of using flexible probabilistic frameworks built without large-scale clinical data sets to generate ranked differential diagnoses that are more accurate that those developed by physicians.

Combining the approaches of this prior work suggests that an appropriate diagnostic support model should rely on easily obtained symptom data, probabilistic frameworks to avoid the need for large-scale data sets, and most importantly, a flexible schema to refine predictions based on provider judgment and the ability to adapt to changes in local prevalence and current diagnostic test performance. To this end, we present a comparison and clinical validation of 3 such quantitative models as well as an ensemble approach to the diagnosis of COVID-19 in ambulatory and acute care settings. We then illustrate how this approach can be employed to help providers optimally reduce diagnostic uncertainty through appropriate diagnostic test choices and update predictions based on local clinical context and test results as that are obtained. Finally, we provide an interactive, online resource to assess COVID-19 infection probability based on user-defined parameters such as local disease prevalence, imaging, and testing performance [17].

Methods

Data Acquisition

National and state-specific confirmed cases of COVID-19 as of July 2, 2020, were acquired from the Center for Systems Science and Engineering at Johns Hopkins University [18]. During our model training, validation, and testing process, we assumed a national SARS-CoV-2 infection prevalence of 11.1% based on the total confirmed count of 5,438,325 in the United States as of August 17, 2020 [18], a population estimate of 328,239,523 [19], and an estimated reporting rate of 14.9% [20-22]. Prevalence and conditional symptom probabilities for diseases in the differential diagnosis were collected from the CDC and literature estimates (Multimedia Appendix 1, Table S1). COVID-19 symptom probabilities were developed primarily from a 393-person consecutive patient series [23] and supplemented by 3 meta-analyses, which included 3062 [24], 49,504 [25], and 53,000 patients [26]. Where conditional symptom probabilities have not been described in the literature, we used a symptom probability of 1.0% based on our assumption that a higher conditional symptom probability would have been discussed in the literature.

To incorporate location and diagnostic test results into risk predictions, we used state-level case figures [18], state-level population data [19], and the estimated reporting rate [20-22] to compute an estimated SARS-CoV-2 infection prevalence for each state. We sourced imaging diagnostic accuracies from existing literature [6,7] and laboratory test accuracies from the Johns Hopkins Center for Health Security. The reverse transcription-polymerase chain reaction (RT-PCR) sensitivity of 70% is based on published estimates [4] that take into account operator dependency and variability in viral load across upper respiratory tract sites [27]. The RT-PCR specificity of 99.8% is based on published data from Abbott Molecular [28]. Antibody test sensitivity and specificity are based on published figures [5] for electro-chemiluminescence immunoassay completed between 0-6 days of infection. We computed likelihood ratios and prevalence-adjusted predictive values based on sensitivity, specificity, and our estimated national COVID-19 prevalence of 11.1% (Table 1).

Table 1. Imaging and laboratory diagnostic test statistics for SARS-CoV-2 infection.

Diagnostic test	Sensitivity (%)	Specificity (%)	Likelihood ratio (%)		Predictive value[a] (%)	
			Positive	Negative	PPV[b]	NPV[c]
Chest radiography	69.0 [6]	50.0[d]	1.4	0.6	14.7	92.8
Computed tomography [7]	97.0	57.3	2.3	0.1	22.1	99.3
RT-PCR[e]	70.0 [4]	97.0 [28]	23.3	0.3	74.5	96.3
Antibody (0-6 days) [5]	65.5	99.8	327.5	0.3	97.6	95.9

[a] Prevalence-adjusted predictive values assume a prevalence of 11.1%.

[b] PPV: positive predictive value.

[c] NPV: negative predictive value.

[d] No published figures available.

[e] RT-PCR: reverse transcription-polymerase chain reaction.

Training

We developed Bayesian inference network (BN) and set-cover (SC) models from published disease prevalence and conditional symptom probabilities (see Multimedia Appendix 1, Table S1). We simulated symptom profiles and diagnoses for 100,000 patients using the published aggregate diagnosis prevalence and conditional symptom probabilities. Of the 100,000 simulated patients, the number of patients assigned to each mutually exclusive diagnosis was proportional to diagnosis prevalence. Within each diagnostic class, we simulated a joint symptom distribution by randomly assigning the presence or absence of each symptom to every patient. The number of patients with a positive symptom within each diagnostic class was proportional to the conditional symptom probability for that symptom and diagnosis. We trained our distance metric learning (DML) and ensemble models on this simulated data.

Study Design

We analyzed consecutive ambulatory and hospitalized patients with COVID-19–compatible syndromes presenting to University of California San Diego Medical Center over 14 days in March and April 2020, with institutional review board approval (#200498). Patients were included if they had a recorded presenting illness including fever or cough, and at least a single NAAT in the electronic health record. Patients were labeled “positive” if they had one or more positive RT-PCR tests and a compatible syndrome or findings on radiographic imaging. Patients were labeled “negative” if they had 2 or more consecutive negative NAAT tests (>72 hours apart) or a single negative RT-PCR and a negative antibody test within 14-21 days of symptom onset. Chart review was performed manually by an infectious disease specialist with an anonymized and blinded data set presented for analysis (see Multimedia Appendix 1 for additional details).

Data Analysis

We calculated the sensitivity, specificity, and prevalence-adjusted accuracy as well as the prevalence-adjusted negative predictive value (NPV) and positive predictive value (PPV) of each model on the clinical test data using standard Wald-type CIs [29]. We estimated the 95% CIs for sensitivity and specificity using Clopper-Pearson exact binomial proportion CIs [29]. We estimated 95% CIs for accuracy using the normal approximation method [29]. For the imaging and laboratory tests, we computed likelihood ratios based on sensitivity and specificity; and prevalence-adjusted predictive values based on sensitivity, specificity, and an assumed national COVID-19 prevalence of 11.1%.

Results

Patient Characteristics

In total, 55 individuals and the presence or absence of 13 symptoms at initial presentation were included in our clinical test data set. Of this, 38 patients (69.1%) were confirmed SARS-CoV-2 infection positive by RT-PCR; 44 subjects were seen via inpatient services, and 11 were seen as outpatients. The majority of subjects (n=43, 78.2%) presented with fever, 63.6% (n=35) with shortness of breath or dyspnea, 54.5% (n=30) with nonproductive cough, 21.8% (n=12) with productive cough, 50.9% (n=28) with fatigue or exhaustion, 9.1% (n=5) with loss of smell, 7.3% (n=4) with sore throat or pharyngalgia, 18.2% (n=10) with body or muscle aches, 16.4% (n=9) with headaches, 16.4% (n=9) with diarrhea, 14.5% (n=8) with nausea, 5.5% (n=3) with vomiting, and 3.6% (n=2) with nasal congestion or rhinorrhea (Table 2; Multimedia Appendix 1, Table S2).

Table 2. Clinical test data set: patient characteristics.

Characteristic	Total, n (%)	SARS-CoV-2 test result	
		Positive, n (%)	Negative, n (%)
Patients	55 (100)	38 (69)	17 (31)
Sex			
Male	27 (49)	21 (55)	6 (35)
Female	28 (51)	17 (45)	11 (65)
Age (years)			
<60	27 (49)	19 (50)	8 (47)
60-70	13 (24)	9 (24)	4 (24)
70-80	9 (16)	5 (13)	4 (24)
>80	6 (11)	5 (13)	1 (6)
Setting			
Inpatient	44 (80)	32 (84)	12 (71)
Outpatient	11 (20)	6 (16)	5 (29)
Symptoms			
Fever	43 (78)	33 (87)	10 (59)
Dyspnea	35 (64)	33 (87)	2 (12)
Dry cough	30 (55)	28 (74)	2 (12)
Productive cough	12 (22)	4 (11)	8 (47)
Fatigue	28 (51)	21 (55)	7 (41)
Loss of smell	5 (9)	5 (13)	0 (0)
Sore throat	4 (7)	1 (3)	3 (18)
Body/muscle aches	10 (18)	8 (21)	2 (12)
Headache	9 (16)	6 (16)	3 (18)
Diarrhea	9 (16)	5 (13)	4 (24)
Nausea	8 (15)	3 (8)	5 (29)
Vomiting	3 (6)	1 (3)	2 (12)
Nasal congestion/rhinorrhea	2 (4)	0 (0)	2 (12)
Comorbidities			
Cancer	9 (16)	5 (13)	4 (24)
Diabetes	12 (22)	10 (26)	2 (12)
Cardiovascular disease	7 (13)	4 (11)	3 (18)
Hypertension	15 (27)	12 (32)	3 (18)

Classification Performance in the Clinical Test Data Set

Base models classified SARS-CoV-2 infection in the clinical test data set with sensitivities and specificities of 81.6% (95% CI 65.7 to 92.3) and 58.8% (95% CI 32.9 to 81.6) for the BN model; 0.0% (95% CI 0.0 to 9.3) and 100.0% (95% CI 80.5 to 100.0) for the SC model; 84.2% (95% CI 68.7 to 94.0) and 64.7% (95% CI 38.3 to 85.8) for the DML model; and 81.6% (95% CI 65.7 to 92.3) and 70.6% (95% CI 44.0 to 89.7) for the ensemble model. The overall accuracy of each of these models was 61.4% (95% CI 48.5 to 74.2) for the BN model; 88.9% (95% CI 80.6 to 97.2) for the SC model; 66.9% (95% CI 54.4 to 79.3) for the DML model; and 71.8% (95% CI 59.9 to 83.7) for the ensemble model. The prevalence-adjusted positive and negative predictive values for each model were 19.9% (95% CI 10.5 to 29.2) and 96.2% (95% CI 92.5 to 100.0) for the BN model; 0.0% and 88.9% for the SC model; 23.0% (95% CI 11.3 to 34.7) and 97.0% (95% CI 93.8 to 100.0) for the DML model; and 25.8% (95% CI 11.4 to 40.2) and 96.8% (95% CI 93.7 to 100.0) for the ensemble model (Table 3).

Table 3. Classification performance on the clinical test data set for the developed base and ensemble models compared to a logistic regression model reported in the literature.

Model	Sensitivity (95% CI) (%)	Specificity (95% CI) (%)	Accuracy[a] (95% CI) (%)	Predictive value[a] (%)	
				PPV[b] (95% CI)	NPV[c] (95% CI)
Bayesian inference network	81.6 (65.7-92.3)	58.8 (32.9-81.6)	61.4 (48.5-74.2)	19.9 (10.5-29.2)	96.2 (92.5-100.0)
Information-theoretic set cover	0.0 (0.0-9.3)	100.0 (80.5-100.0)	88.9 (80.6-97.2)	0.0	88.9
Distance metric learning	84.2 (68.7-94.0)	64.7 (38.3-85.8)	66.9 (54.4-79.3)	23.0 (11.3-34.7)	97.0 (93.8-100.0)
Multinomial logistic regression ensemble	81.6 (65.7-92.3)	70.6 (44.0-89.7)	71.8 (59.9-83.7)	25.8 (11.4-40.2)	96.8 (93.7-100.0)
Logistic regression (Menni et al [14])	15.8 (6.0-31.3)	100.0 (80.5-100.0)	90.6 (82.9-98.3)	100.0	90.5 (88.7-92.2)

[a] Prevalence-adjusted metrics assume a COVID-19 prevalence of 11.1%.

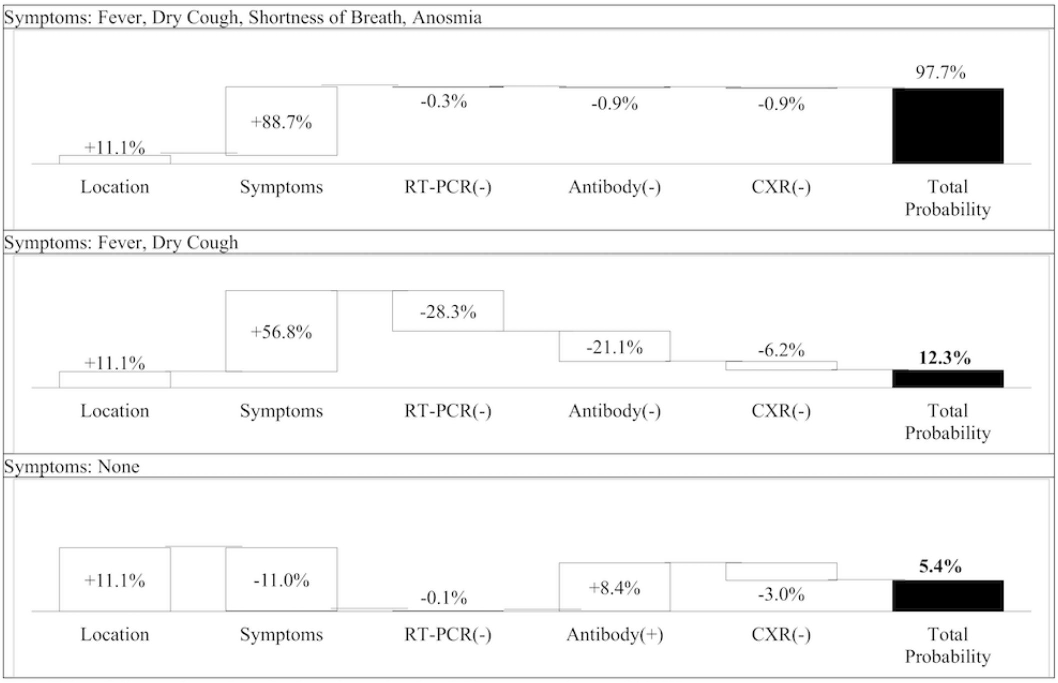
[b] PPV: positive predictive value.

[c] NPV: negative predictive value.

Incorporation of Location and Diagnostic Test Sequences

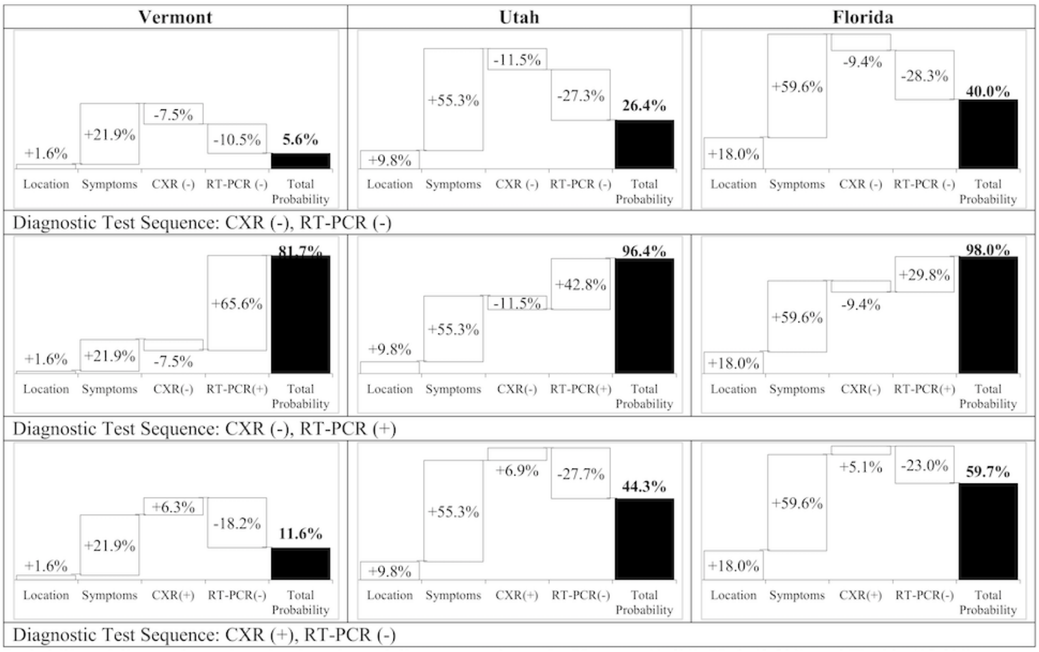
We then employed the BN model to evaluate 3 hypothetical patients with 3 different presentations: (1) fever, dry cough, shortness of breath, and anosmia; (2) fever and dry cough; and (3) asymptomatic. We assumed all of these patients presented for care in an area with a local disease prevalence equivalent to the national disease prevalence of 11.1%. For patient 1, we simulated a clinically plausible imaging and test result sequence of negative RT-PCR, negative antibody, and negative CXR. The probability of a COVID-19 diagnosis following symptom collection was 99.8%. Despite negative test results, residual risk due to local disease prevalence and symptoms remained at 97.7%. The change in diagnosis probability, or the reduction in diagnostic uncertainty, was only 2.1% following all 3 negative tests. For patient 2, we simulated the same negative test sequence. In this scenario, the combination of negative test results with nonspecific symptom information resulted in a decrease in residual risk to 12.3%. The reduction in diagnostic uncertainty due to test results was 55.6%, primarily due to negative RT-PCR and negative antibody test results. The negative CXR provided less information as the reduction in diagnostic uncertainty following negative RT-PCR and antibody tests was only 6.2%. For patient 3, we simulated an imaging and test result sequence of negative RT-PCR, positive antibody, and negative CXR. The negative RT-PCR test reduced disease probability by only 0.1%, and the positive antibody test increased the probability of a COVID-19 diagnosis by 8.4%. The CXR results reduced diagnostic uncertainty by 3.0% (Figure 1).

Figure 1. Probability of SARS-CoV-2 infection for common patient presentations and clinical test sequences. Probability of disease due to location is the estimated national disease prevalence of 11.1%. RT-PCR: reverse transcription-polymerase chain reaction; CXR: chest radiography.



To illustrate the dependence of risk assessment on local disease prevalence, we simulated a patient with symptoms of only fever and dry cough presenting in 3 locations with significantly different COVID-19 prevalence estimates: Vermont with an estimated statewide prevalence of 1.6%, Utah with an estimated statewide prevalence of 9.8%, and Florida with an estimated statewide prevalence of 18.0% at the time of the simulation. We combined results from 3 common test sequences with our BN pretest probabilities to compute location-dependent risk trajectories. The test sequences included: (1) negative CXR and negative RT-PCR; (2) negative CXR and positive RT-PCR; and (3) positive CXR and negative RT-PCR. Our results indicate that for a pauci-symptomatic patient presenting with identical symptoms in states with significantly different disease prevalence, the posttest probabilities of SARS-CoV-2 infection following common diagnostic test sequences demonstrate marked variation. Moreover, changes in diagnostic probability or reductions in diagnostic uncertainty are highly context and test dependent (Figure 2).

Figure 2. Impact of patient location and diagnostic test results on probability of SARS-CoV-2 infection. Prior probability of disease due to location is the estimated disease prevalence for Vermont (1.6%), Utah (9.8%), and Florida (18.0%). Incremental probability due to symptoms assumes the patient presents with only fever and dry cough. RT-PCR: reverse transcription-polymerase chain reaction; CXR: chest radiography.



Discussion

Principal Findings

Our results suggest simple computable models that quantify patient risk of SARS-CoV-2 infection based on key elements of the clinical case can reduce diagnostic uncertainty for providers attempting to rule in or rule out disease with limited or conflicting information.

Building on work by Chishti et al [16], we chose probabilistic models considering the scarcity of detailed, individual patient data and to take advantage of the depth of published literature on aggregate symptom probabilities. Clinicians are trained in evidence-based medicine, and Bayesian reasoning provides a natural framework to understand the impact of incremental information on diagnosis probabilities. Our approaches to making stepwise diagnostic assessments with incremental information mimic clinical workflows and reflect the need for transparency and accommodation of new information critical to clinical decision making. As in Menni et al [14], we chose clinical indicators that would be easily obtained by patients and providers as well as predictive models that are easily computed and transparent to all users. While other machine learning approaches, such as generative adversarial networks, transfer learning, n-shot learning, and prototypical networks, are also robust for limited data, these methods can be opaque and inaccessible to providers and may be inflexible and fragile in an evolving clinical context.

Our most simple model, the Bayesian inference network, is transparent, easily interpreted, and highly modifiable depending on the user’s prior beliefs about location-based prevalence, conditional symptom probabilities, and imaging and laboratory test accuracy. Clinicians, educated in evidence-based medicine and often familiar with Bayesian decision making in diagnostic testing, are ideal users of this model. By developing base models that do not require access to large amounts of patient-level data and can accommodate changes in local provider beliefs and new sources of information, we alert physicians to the utility of using Bayesian reasoning to not only combine multiple data streams in order to make more informed diagnostic decisions but also to guide decisions about use of imaging and testing that will most effectively reduce diagnostic uncertainty.

Limitations

Our study has limitations. First, we used simulated patient data based on prevalence and conditional symptom probabilities to train and validate our DML and ensemble models that biased the ensemble model to heavily weight the DML model predictions. Second, the number of patients in our clinical test data set was relatively small, and this data set was enriched for SARS-CoV-2–positive patients due to the cancellation of all elective procedures and the use of telemedicine for almost all patient visits during the study period, leaving clinics and hospitals open primarily for COVID-19 patients and the acutely ill. Third, 80% of the patients in our clinical test data set were from inpatient services, potentially biasing model accuracy by disease severity. Fourth, we chose as a reference standard the RT-PCR test results for SARS-CoV-2 infection despite outstanding questions about false negative rates in NAAT tests due to operator dependency and patient-level differences in viral loads across upper respiratory tract sites [4,27].

Conclusions

Overall, we found that the Bayesian inference network, the metric learning model, and ensemble models trained and validated on a simulated patient data set had sensitivities (81.6%-84.2%) and specificities (58.8%-70.6%) for discriminating between COVID-19 infection and other potential diagnoses in real clinical settings. These models had higher sensitivities than reported for most commonly used diagnostics, and model specificities were higher than those of both imaging modalities. For purposes of comparison, the logistic regression model proposed by Menni et al [14], when applied to our clinical test data set, had a sensitivity of 15.8% and a specificity of 100.0%. Finally, our BN model shows that information acquired by imaging and testing choices is highly dependent on location and symptoms, and emphasizes the utility of a quantitative framework to guide clinical decision making in rapidly changing local environments with potentially unreliable diagnostic tests.

Acknowledgments

We would like to thank Mary R Mulcare, MD, at the Weill Cornell School of Medicine; Dana R Levin, MD, MPH, at the Columbia University College of Physicians and Surgeons; and Daniel L Wang, MD, at the University of California San Diego School of Medicine, for scientific guidance, help with data preparation, and manuscript comments.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental materials.
[DOCX File , 315 KB - jmir_v22i12e24478_app1.docx]

References

1. Song J, Yun J, Noh J, Cheong H, Kim W. Covid-19 in South Korea — Challenges of Subclinical Manifestations. N Engl J Med 2020 May 07;382(19):1858-1859. [doi: 10.1056/nejmc2001801]
2. Coronavirus Disease 2019 (COVID-19): Overview of Testing for SARS-CoV-2 (COVID-19). Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing-overview.html [accessed 2020-09-15]
3. Bordi L, Nicastri E, Scorzolini L, Di Caro A, Capobianchi MR, Castilletti C, et al. Differential diagnosis of illness in patients under investigation for the novel coronavirus (SARS-CoV-2), Italy, February 2020. Euro Surveill 2020 Feb;25(8):2000170 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.8.2000170] [Medline: 32127123]
4. Woloshin S, Patel N, Kesselheim AS. False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications. N Engl J Med 2020 Aug 06;383(6):e38. [doi: 10.1056/nejmp2015897]
5. Global Progress on COVID-19 Serology-Based Testing. Johns Hopkins Center for Health Security. 2020. URL: https://www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html [accessed 2020-09-15]
6. Wong H, Lam H, Fong A, Leung ST, Chin TWY, Lo CSY, et al. Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. Radiology 2020 Aug;296(2):E72-E78 [FREE Full text] [doi: 10.1148/radiol.2020201160] [Medline: 32216717]
7. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology 2020 Aug;296(2):E32-E40 [FREE Full text] [doi: 10.1148/radiol.2020200642] [Medline: 32101510]
8. Grinfeld J, Nangalia J, Baxter EJ, Wedge DC, Angelopoulos N, Cantrill R, et al. Classification and Personalized Prognosis in Myeloproliferative Neoplasms. N Engl J Med 2018 Oct 11;379(15):1416-1430. [doi: 10.1056/nejmoa1716614]
9. Palmerini T, Benedetto U, Bacchi-Reggiani L, Riva DD, Biondi-Zoccai G, Feres F, et al. Mortality in patients treated with extended duration dual antiplatelet therapy after drug-eluting stent implantation: a pairwise and Bayesian network meta-analysis of randomised trials. The Lancet 2015 Jun;385(9985):2371-2382. [doi: 10.1016/s0140-6736(15)60263-x]
10. Milea D, Najjar RP, Jiang Z, Ting D, Vasseneix C, Xu X, et al. Artificial Intelligence to Detect Papilledema from Ocular Fundus Photographs. N Engl J Med 2020 Apr 30;382(18):1687-1695. [doi: 10.1056/nejmoa1917130]

11. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Jan 25;542(7639):115-118. [doi: 10.1038/nature21056]
12. Liu Y, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. JAMA 2019 Nov 12;322(18):1806-1816. [doi: 10.1001/jama.2019.16489] [Medline: 31714992]
13. Rosenbaum L, Lamas D. Cents and sensitivity--teaching physicians to think about costs. N Engl J Med 2012 Jul 12;367(2):99-101. [doi: 10.1056/NEJMp1205634] [Medline: 22784112]
14. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul 11;26(7):1037-1040. [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
15. Ahsan MM, E. Alam T, Trafalis T, Huebner P. Deep MLP-CNN Model Using Mixed-Data to Distinguish between COVID-19 and Non-COVID-19 Patients. Symmetry 2020 Sep 16;12(9):1526. [doi: 10.3390/sym12091526]
16. Chishti S, Jaggi KR, Saini A, Agarwal G, Ranjan A. Artificial Intelligence-Based Differential Diagnosis: Development and Validation of a Probabilistic Model to Address Lack of Large-Scale Clinical Datasets. J Med Internet Res 2020 Apr 28;22(4):e17550 [FREE Full text] [doi: 10.2196/17550] [Medline: 32343256]
17. COVID-19 Infection Probability Assessment. 2020. URL: http://covid-calc.herokuapp.com/ [accessed 2020-09-15]
18. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
19. State Population Totals and Components of Change: 2010-2019. United States Census Bureau. 2019. URL: https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html [accessed 2020-09-15]
20. Commercial Laboratory Seroprevalence Data. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/commercial-lab-surveys.html [accessed 2020-09-15]
21. Havers F, Reed C, Lim T, Montgomery JM, Klena JD, Hall AJ, et al. Seroprevalence of Antibodies to SARS-CoV-2 in Six Sites in the United States, March 23-May 3, 2020. medRxiv Preprint posted online June 26, 2020. [doi: 10.1101/2020.06.25.20140384]
22. Pei S, Shaman J. Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US. medRxiv Preprint posted online March 27, 2020. [doi: 10.1101/2020.03.21.20040303]
23. Goyal P, Choi JJ, Pinheiro LC, Schenck EJ, Chen R, Jabri A, et al. Clinical Characteristics of Covid-19 in New York City. N Engl J Med 2020 Jun 11;382(24):2372-2374. [doi: 10.1056/nejmc2010419]
24. Zhu J, Ji P, Pang J, Zhong Z, Li H, He C, et al. Clinical characteristics of 3062 COVID-19 patients: A meta-analysis. J Med Virol 2020 Oct 24;92(10):1902-1914 [FREE Full text] [doi: 10.1002/jmv.25884] [Medline: 32293716]
25. Heydari K, Rismantab S, Shamshirian A, Lotfi P, Shadmehri N, Houshmand P, et al. Clinical and Paraclinical Characteristics of COVID-19 patients: A systematic review and meta-analysis. medRxiv Preprint posted online March 30, 2020. [doi: 10.1101/2020.03.26.20044057]
26. Ma C, Gu J, Hou P, Zhang L, Bai Y, Guo Z, et al. Incidence, clinical characteristics and prognostic factor of patients with COVID-19: a systematic review and meta-analysis. medRxiv Preprint posted online March 20, 2020.
27. Yu F, Yan L, Wang N, Yang S, Wang L, Tang Y, et al. Quantitative Detection and Viral Load Analysis of SARS-CoV-2 in Infected Patients. Clin Infect Dis 2020 Jul 28;71(15):793-798 [FREE Full text] [doi: 10.1093/cid/ciaa345] [Medline: 32221523]
28. Current Molecular and Antigen Tests with FDA EUA Status. Johns Hopkins Center for Health Security. 2020. URL: https://www.centerforhealthsecurity.org/resources/COVID-19/molecular-based-tests/current-molecular-and-antigen-tests.html [accessed 2020-09-15]
29. Mercaldo ND, Lau KF, Zhou XH. Confidence intervals for predictive values with an emphasis to case-control studies. Stat Med 2007 May 10;26(10):2170-2183. [doi: 10.1002/sim.2677] [Medline: 16927452]

Abbreviations

BN: Bayesian network
CDC: Centers for Disease Control and Prevention
CDSS: clinical decision support systems
CT: computed tomography
CXR: chest radiography
DML: distance metric learning
NAAT: nucleic acid amplification test
NPV: negative predictive value
PPV: positive predictive value
RT-PCR: reverse transcription-polymerase chain reaction
SC: set-cover model

Edited by G Eysenbach, R Kukafka; submitted 02.10.20; peer-reviewed by M Majurul Ahsan, N Agarwal; comments to author 11.10.20; revised version received 13.10.20; accepted 02.11.20; published 16.12.20.

Please cite as:
D'Ambrosia C, Christensen H, Aronoff-Spencer E
Computing SARS-CoV-2 Infection Risk From Symptoms, Imaging, and Test Data: Diagnostic Model Development
J Med Internet Res 2020;22(12):e24478
URL: http://www.jmir.org/2020/12/e24478/
doi:10.2196/24478
PMID:33301417

Original Paper

Subjective Well-Being of Chinese Sina Weibo Users in Residential Lockdown During the COVID-19 Pandemic: Machine Learning Analysis

Yilin Wang[1,2,3], BSc; Peijing Wu[1,2,4], BSc; Xiaoqian Liu[1,2], DPhil; Sijia Li[1,2], BSc; Tingshao Zhu[1,2], DPhil; Nan Zhao[1,2], DPhil

[1]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China
[3]Department of Psychology, Nankai University, Tianjin, China
[4]Beijing University of Posts and Telecommunications, Beijing, China

Corresponding Author:
Nan Zhao, DPhil
Key Laboratory of Behavioral Science
Institute of Psychology
Chinese Academy of Sciences
16 Lincui Road
Chaoyang District
Beijing,
China
Phone: 86 010 64101377
Email: zhaonan@psych.ac.cn

Abstract

Background: During the COVID-19 pandemic, residential lockdowns were implemented in numerous cities in China to contain the rapid spread of the disease. Although these stringent regulations effectively slowed the spread of COVID-19, they may have posed challenges to the well-being of residents.

Objective: This study aims to explore the effects of residential lockdown on the subjective well-being (SWB) of individuals in China during the COVID-19 pandemic.

Methods: The sample consisted of 1790 Sina Weibo users who were residents of cities that imposed residential lockdowns, of which 1310 users (73.18%) were female, and 3580 users who were residents of cities that were not locked down (gender-matched with the 1790 lockdown residents). In both the lockdown and nonlockdown groups, we calculated SWB indicators during the 2 weeks before and after the enforcement date of the residential lockdown using individuals' original posts on Sina Weibo. SWB was calculated via online ecological recognition, which is based on established machine learning predictive models.

Results: The interactions of time (before the residential lockdown or after the residential lockdown) × area (lockdown or nonlockdown) in the integral analysis (N=5370) showed that after the residential lockdown, compared with the nonlockdown group, the lockdown group scored lower in some negative SWB indicators, including somatization ($F_{1,5368}$=13.593, P<.001) and paranoid ideation ($F_{1,5368}$=14.333, P<.001). The interactions of time (before the residential lockdown or after the residential lockdown) × area (developed or underdeveloped) in the comparison of residential lockdown areas with different levels of economic development (N=1790) indicated that the SWB of residents in underdeveloped areas showed no significant change after the residential lockdown (P>.05), while that of residents in developed areas changed.

Conclusions: These findings increase our understanding of the psychological impact and cost of residential lockdown during an epidemic. The more negative changes in the SWB of residents in developed areas imply a greater need for psychological intervention under residential lockdown in such areas.

(J Med Internet Res 2020;22(12):e24775) doi:10.2196/24775

KEYWORDS

COVID-19; residential lockdown; subjective well-being; online ecological recognition

Introduction

Since the outbreak of COVID-19, therapeutic options to treat the disease have been limited, although research to produce an effective and safe vaccine is quickly progressing. Varying degrees of travel limitations on citizens, such as residential lockdowns, were enforced across China to contain the epidemic. Residential lockdown is a type of home quarantine during which residents are not allowed to go out unnecessarily and each family can only appoint a relatively fixed family member to purchase living supplies every 2-3 days. These vigorous actions played a key role in containing the spread of COVID-19 [1,2]; however, they have also had some effects on residents' mental health, including their subjective well-being (SWB) [3,4]. In psychology, the concepts of well-being, mental health, and happiness are often used as synonyms [5]. The World Health Organization (WHO) defined mental health as a "state of well-being whereby individuals recognize their abilities, are able to cope with the normal stresses of life, work productively and fruitfully, and make a contribution to their communities" [6].

As a comprehensive psychological indicator, SWB describes an individual's subjective evaluation of their living quality and happiness [7]. It focuses on personal experience in a specific social context to reflect individuals' mental health [8-11]. As a social group experience [11], SWB has also become a surveying mark of government action, governance capability, and social development trends [12,13]. During the COVID-19 crisis, the loss or significant reduction of SWB may not only increase the incidence of panic disorder, anxiety, depression, and other mental health issues [14] but also cause a series of problems at the social level, such as social hindrance, unstable social environment, and decline of trust in the government [15]. To adequately and effectively prepare for the psychological impact and cost of residential lockdown (eg, targeted psychological crisis intervention), it is necessary to examine the mental health of residents from the SWB perspective.

At present, the assessment of positive SWB indicators mainly relies on self-report inventories such as the Psychological Well-Being (PWB) scale [16]. In addition, because mental health is the crucial element or is even considered to be a synonym of SWB [5,6], some psychiatric status rating scales are widely used to assess negative SWB indicators [17,18], such as the Symptom Checklist 90 (SCL-90), with good reliability and validity [19-21].

However, it is almost impossible to conduct a SWB survey measuring the immediate psychological impact of residential lockdown without delay, as it is difficult to anticipate whether and when a residential lockdown will start, and a questionnaire cannot be used for ecological momentary assessment [22,23]. On the other hand, retrospective measurement is also unsuitable for this aim because individuals' recall bias may result in inability to accurately trace back to the original baseline state [24]. Facing these challenges in conducting an accurate pre-post comparative analysis during unexpected residential lockdown, it is necessary to develop a new method to capture the SWB status of a large number of individuals in a timely fashion.

With the rise of interdisciplinary research, information technology and computing models are increasingly being used in the evaluation of psychological characteristics, and the popularity of web-based social media provides a new platform for such research. Information collected from these social media platforms can clarify the public's response and can be mapped to psychological indices [25]. For example, Kosinski et al [26] used digital records on Facebook to automatically and accurately predict users' personal attributes. Qiu et al [27] also showed that users' personalities manifested on web-based social media. In China, Sina Weibo is a leading social media platform; by March 2020, it comprised more than 550 million monthly active users. By using machine learning–trained predictive models, users' ecological behavior on Sina Weibo can be tracked to identify their psychological traits, such as their mental health status [28]. Based on web-based big data, these prediction models provided a rationale for our research. According to previous studies, online ecological recognition (OER) can be defined as the method of acquiring users' psychological characteristics by machine learning models based on ecological web-based social media data [29,30]. Therefore, in this study, to effectively examine the impact of the residential lockdown on SWB, we conducted OER on publicly available Sina Weibo data to identify users' SWB with high ecological validity on a large scale.

In the present study, we aimed to explore the impact of the residential lockdown on SWB of individuals during the COVID-19 pandemic. To examine individuals' SWB before and after the residential lockdown, we used data from posts by active Sina Weibo users to calculate SWB indicators through OER based on trained predictive models [5]. The main purpose of this study was to obtain empirical evidence to provide deep insight into the psychological impact and cost of the residential lockdown. This will in turn form the basis for a concrete plan to develop and implement targeted intervention policies to cope with the COVID-19 pandemic efficiently and effectively.

Methods

Sampling

We searched for the following keywords on Baidu (baidu.com): *residential lockdown* (居民出行管控) + *city* or *211 measure* (211举措; the abbreviation of the residential lockdown policy) + *city*. After locating the cities that had issued residential lockdowns from January 20 to February 20, 2020, we finally obtained 17 prefecture-level cities in which local governments implemented the same residential lockdown policy. Then, we identified the specific time period for the implementation of the residential lockdown from the announcement released by the local COVID-19 Prevention and Control Headquarters, and we took the enforcement date as the demarcation point of this study. The residential lockdown cities and the enforcement and ending dates of their residential lockdowns are shown in Table 1. Given that the lockdown enforcement durations of Ningbo, Taizhou, and Yiwu were all <2 weeks, we excluded these three residential lockdown cities from the subsequent analysis.

Table 1. Residential lockdown cities and their enforcement and ending dates (in 2020).

Residential lockdown city	Province	Enforcement date	Ending date
Bengbu	Anhui	February 3	March 22
Ezhou	Hubei	February 4	March 25
Fangchenggang	Guangxi Zhuang Autonomous Region	February 2	March 6
Fuyang	Anhui	February 5	March 9
Guigang	Guangxi Zhuang Autonomous Region	February 3	February 24
Huaibei	Anhui	February 3	March 20
Huanggang	Hubei	February 1	March 22
Huangshi	Hubei	February 17	March 23
Nanyang	Henan	February 4	March 11
Ningbo[a]	Zhejiang	February 4	February 10
Songyuan	Jilin	February 4	March 20
Taizhou[a]	Zhejiang	February 2	February 13
Wenzhou	Zhejiang	February 1	February 19
Wuhan	Hubei	February 11	April 8
Xiaogan	Hubei	February 14	March 14
Yiwu[a]	Zhejiang	February 4	February 12
Zhongshan	Guangdong	February 7	March 27

[a] Ningbo, Taizhou, and Yiwu were not included in subsequent analyses because the enforcement duration in these cities was less than 2 weeks.

Participants and Data Collection

The sample in this study was obtained from the original Sina Weibo data pool [30] containing more than 1.16 million active users. The retrieved data included users' account profiles and posts. Privacy was strictly protected during this process, referring to the ethical principles listed by Kosinski et al [31]. This study was approved by the Review Board of Institute of Psychology, Chinese Academy of Sciences (ethical code H15009).

From the data pool, we selected Sina Weibo users who met the following criteria: (1) Published at least one original Weibo post per day on average within 2 weeks before and after the corresponding enforcement date of the residential lockdown (see Table 2); (2) authentication type was no-institutional, such as individual user; (3) the geolocation of the user's account profile was in China, not "overseas" or "others."

Lockdown Group

We combined all the residential lockdown cities mentioned in Table 1 except Ningbo, Taizhou, and Yiwu as the lockdown group. Then, we downloaded each user's original posts during the 2 weeks before and after the enforcement date of the residential lockdown in their city of residence for analysis in this study.

Nonlockdown Group

We combined users who did not live in the abovementioned 17 residential lockdown cities as the entire nonlockdown sample group. Because cities under residential lockdown are a minority among all the cities in China, to increase the generalizability of the results and reduce sampling error, we built the nonlockdown group at twice the sample size of the lockdown group in each of the 14 residential lockdown cities by random sampling without replacement of users in the whole nonlockdown sample group. The male-female ratio of the nonlockdown group was same as that of the lockdown group. Finally, we downloaded each user's original posts during the 2 weeks before and after the enforcement date from the paired residential lockdown city for analysis in this study.

The final sample contained 5370 users, among whom the male-female ratio was 1:2.7 (see Table 2 for details).

Table 2. Numbers of users in each group in the final sample and corresponding time periods of data extraction (N=5370).

Residential lockdown city	Time period (in 2020)		Lockdown group	Nonlockdown group
	Before the residential lockdown	After the residential lockdown		
Bengbu	January 21 to February 3	February 4-17	25	50
Ezhou	January 21 to February 3	February 4-17	16	32
Fangchenggang	January 19 to February 1	February 2-15	18	36
Fuyang	January 22 to February 4	February 5-18	14	28
Guigang	January 20 to February 2	February 3-16	32	64
Huaibei	January 20 to February 2	February 3-16	25	50
Huanggang	January 18-31	February 1-14	28	56
Huangshi	February 3-16	February 17 to March 1	11	22
Nanyang	January 21 to February 3	February 4-17	42	84
Songyuan	January 21 to February 3	February 4-17	16	32
Wenzhou	January 19 to February 1	February 2-15	235	470
Wuhan	January 28 to February 10	February 11-24	1084	2168
Xiaogan	January 31 to February 13	February 14-27	10	20
Zhongshan	January 24 to February 6	February 7-20	234	468

Compartmentalization by Economic Development Level

SWB is influenced by various factors in the long term [32-35], and there is an inevitable connection between the level of economic development and SWB. Clark and Senik [36] reported that economic development was generally identified with growth in gross domestic product (GDP) per capita, which improved residents' welfare [37]. Therefore, differentiating the lockdown group by GDP per capita could clarify the impact of the residential lockdown under different economic development conditions.

We took the GDP per capita published by the Municipal Bureau of Statistics of each residential lockdown city in 2017 as the indicator of economic development level. Using between-groups linkage and taking the squared Euclidean distance as the measurement standard, we applied a hierarchical clustering analysis on the 14 residential lockdown cities included in our study. The dendrogram of the whole process of clustering indicated that when the 14 residential lockdown cities were divided into 2 categories, the distance between the categories was relatively large and the characteristics of each category were relatively prominent; therefore, the categories could be readily defined. We applied an additional hierarchical clustering analysis with a cluster number of 2 on the 14 residential lockdown cities. The results suggested that the first category included Wuhan, Ezhou, Zhongshan, and Fangchenggang, and the second category included the remaining 10 residential lockdown cities. Based on this result, in this study, we grouped Wuhan, Ezhou, Zhongshan, and Fangchenggang into developed areas with residential lockdowns and grouped the remaining 10 residential lockdown cities into underdeveloped areas with residential lockdowns. All analyses were performed using SPSS Statistics 26 (IBM Corporation).

Measurement of SWB and Procedure

We applied OER by employing established machine learning models [28,38] to calculate the SWB indicators of the users in the final sample before and after the corresponding enforcement date of the residential lockdown. First, we used the TextMind system [39], a Chinese language psychological analysis system, to extract psycholinguistic features from active Sina Weibo users' original posts. This system first segmented the users' original Weibo posts and extracted independent and linguistically labeled words [40]; then, it used the Simplified Chinese Language Inquiry and Word Count (SCLIWC) to determine the word frequency statistics [41].

Next, we used these extracted psycholinguistic features as the input of established machine learning models to calculate the SWB indicators. The SWB indicators in our research included both positive and negative indicators, which are depicted in Table 3. Because the SCL-90 is also widely used in the assessment of negative SWB indicators [17], we used factors of the SCL-90 as our negative SWB indicators.

Table 3. Positive and negative SWB indicators.

SWB[a] indicator	Definition
Positive[b]	
Self-acceptance	Possession of a positive attitude toward the self
Environmental mastery	Sense of mastery and competence in managing the environment
Positive relations	Warm, satisfying, trusting relationships with others
Purpose in life	Goals in life and a sense of directedness
Personal growth	Feeling of continued development
Autonomy	Self-determining and independent
Negative[c]	
Somatization	Distress arising from perceptions of bodily disfunction
Obsessive-compulsive	Thoughts, impulses, and actions that are experienced as unremitting and irresistible by the individual, but are of an ego-alien or unwanted nature
Interpersonal sensitivity	Feelings of personal inadequacy and inferiority, particularly in comparison to other persons
Depression	Withdrawal of life interest, lack of motivation, and loss of vital energy
Anxiety	Restlessness, nervousness, and tension
Hostility	Thoughts, feelings, or actions that are characteristics of the negative affect state of anger
Phobic anxiety	Persistent fear response to a specific person, place, object, or situation which is characterized as being irrational and disproportionate to the stimulus, and which leads to avoidance or escape behavior
Paranoid Ideation	Paranoid behavior fundamentally as a disordered mode of thinking
Psychoticism	Withdrawn, isolated, schizoid lifestyle, and first-rank symptoms of schizophrenia.

[a] SWB: subjective well-being.

[b] Source: Psychological Well-Being scale [42].

[c] Source: Symptom Checklist 90 [19].

We used Li, Hao, Bai, and Zhu's machine learning model [28] to calculate the positive SWB indicators. To establish this predictive model, 1785 active Sina Weibo users were invited to complete Scales of PWB on the internet [42] and were asked for approval to access their Weibo data. Then, we used multivariate adaptive regression splines and other algorithms to fit the relationship between the psycholinguistic features and self-report scores, and the evaluation was conducted by 10-fold cross-validation. To predict the negative SWB indicators, we used Hao et al's [38] machine-learning model to predict them. To establish this model, we used Weibo data of 448 active Sina Weibo users and their SCL-90 scores [19]. We used a linear regression algorithm to fit the relationship between the psycholinguistic features and self-report scores and conducted 10-fold cross-validation. For most SWB indicators, the Pearson correlation coefficient between the predicted and self-report scores achieved a moderate level (≥.30) [28,38]. For more details about the established SWB models, please see [28,38].

To explore the overall impact of the residential lockdown on SWB, we conducted 2 (times: before the residential lockdown, after the residential lockdown) × 2 (areas: lockdown, nonlockdown) repeated measures analysis of variance (RM ANOVA) on all our samples. In this analysis, time was the within-subject factor, while area was the between-subject factor. Moreover, to explore the impact of the residential lockdown on the SWB of residents at different economic development levels, we applied 2 (times: before the residential lockdown, after the residential lockdown) × 2 (areas: developed, underdeveloped) RM ANOVA on the residential lockdown cities. All these analyses were performed using SPSS 26.

Results

Demographics

Among the 5370 active Sina Weibo users, 1790 were in the lockdown group (1310 female, 73.18%) and 3580 were in the non-lockdown group (gender matched with the lockdown group). According to the birthdates the users registered in their account profiles (1160/5370, 21.60%), the users' ages ranged from 21 to 83 years, with a median age of 29.95 (SD 6.30) years. Table 4 features the demographic profiles of these users.

Table 4. Demographic characteristics of users in the lockdown group and nonlockdown group (N=5370), n (%).

Characteristic	Developed areas with residential lockdown (n=1352)	Underdeveloped areas with residential lockdown (n=438)	Lockdown group (n=1790)	Nonlockdown group (n=3580)
Gender				
Male	348 (25.74)	132 (30.14)	480 (26.82)	960 (26.82)
Female	1004 (74.26)	306 (69.86)	1310 (73.18)	2620 (73.18)
Age (years)				
18-30	198 (14.64)	57 (13.01)	255 (14.25)	464 (12.96)
31-40	92 (6.80)	36 (8.22)	128 (7.15)	249 (6.96)
≥41	11 (0.81)	7 (1.60)	18 (1.01)	46 (1.28)
Missing data	1051 (77.74)	338 (77.17)	1389 (77.60)	2821 (78.80)

All Residential Lockdown Cities

In this analysis, we compared the SWB scores of residents of different areas across the country during different periods. As shown in Table 5, there were significant (including marginally significant) interactions of area and time on 2 negative SWB indicators.

Table 5. Repeated measures analysis of variance on subjective well-being across China.

Indicator	Residential lockdown area, mean (SD)				Area		Time		Area × Time	
	Lockdown group (n=1790)		Nonlockdown group (n=3580)		$F_{1,5368}$	*P* value	$F_{1,5368}$	*P* value	$F_{1,5368}$	*P* value
	T-before[a]	T-after[b]	T-before	T-after						
Somatization	8.38 (2.94)	8.64 (3.04)	8.56 (3.25)	8.98 (3.30)	10.457	.001	56.414	<.001	3.430	.06
Paranoid Ideation	5.46 (1.46)	5.61 (1.45)	5.58 (1.85)	5.82 (2.11)	12.418	<.001	66.862	<.001	3.405	.07

[a] T-before: the period before the residential lockdown.

[b] T-after: the period after the residential lockdown.

Somatization

The main effect of the area was significant ($F_{1,5368}$=10.457, P=.001, η_p^2=0.002). The main effect of time was significant ($F_{1,5368}$=56.414, P<.001, η_p^2=0.010. The interaction of area and time was marginally significant ($F_{1,5368}$=3.430, P=.06, η_p^2=0.001. After simple effect analysis, our findings were as follows:

Area: In the T-before period, there was no significant difference between the lockdown group and the nonlockdown group; in the T-after period, the somatization of residents in the lockdown group was significantly lower than that of residents in the non-lockdown group ($F_{1,5368}$=13.593, P<.001, η_p^2=0.003).

Time: After the implementation of the residential lockdown, somatization of the residents in both groups increased significantly (lockdown group: $F_{1,5368}$=12.009, P=.001, η_p^2=0.002; nonlockdown group: $F_{1,5368}$=65.748, P<.001, η_p^2=0.012).

Paranoid Ideation

The main effect of area was significant ($F_{1,5368}$=12.418, P<.001, η_p^2=0.002). The main effect of time was significant ($F_{1,5368}$=66.862, P<.001, η_p^2=0.012). The interaction of area and time was marginally significant ($F_{1,5368}$=3.405, P=.07, η_p^2=0.001). After simple effect analysis, our findings were as follows:

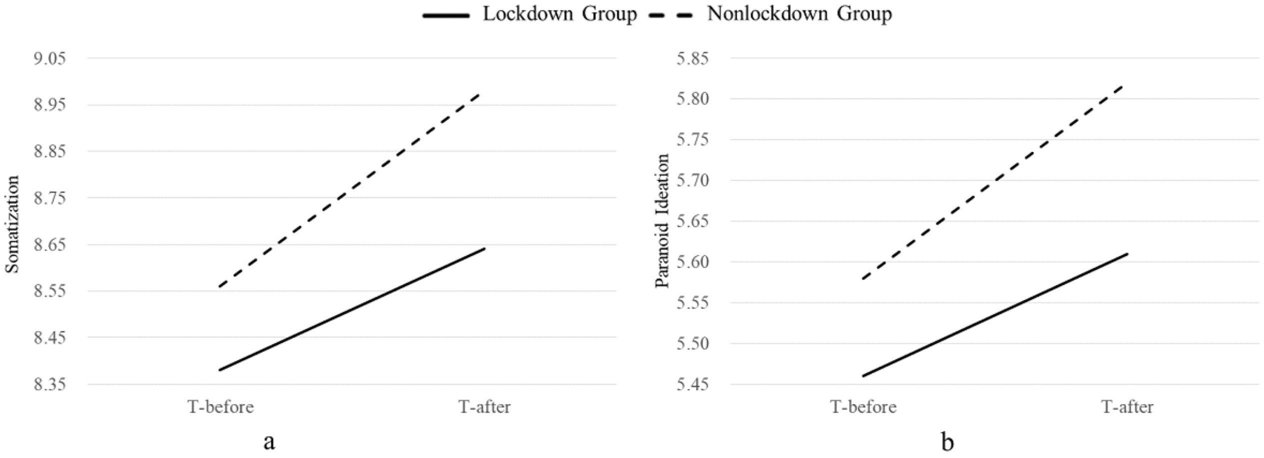
Area: In both periods, the paranoid ideation of residents in the lockdown group was significantly lower than that of residents in the nonlockdown group (T-before: $F_{1,5368}$=6.008, P=.01, η_p^2=0.001; T-after: $F_{1,5368}$=14.333, P<.001, η_p^2=0.003).

Time: After the implementation of the residential lockdown, the paranoid ideation of residents in both groups increased significantly (lockdown group: $F_{1,5368}$=15.033, P<.001, η_p^2=0.003; non-lockdown group: $F_{1,5368}$=75.334, P<.001, η_p^2=0.014).

However, for self-acceptance, environmental mastery, positive relations, purpose in life, personal growth, autonomy, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, and psychoticism, the interactions of area and time were not significant (all *P* values >.10).

The interactions of somatization and paranoid ideation are presented in Figure 1.

Figure 1. Interactions of area and time on (a) somatization and (b) paranoid ideation across China. T-after: the period after the residential lockdown; T-before: the period before the residential lockdown.



Residential Lockdown Areas at Different Economic Development Levels

According to the clustering analysis above, we grouped Wuhan, Ezhou, Zhongshan, and Fangchenggang into developed areas with residential lockdowns, and we grouped the remaining 10 residential lockdown cities into underdeveloped areas with residential lockdowns. To further explore the different effects of residential lockdown on the SWB of residents of areas with different economic development, we compared the SWB scores of residents of developed areas and underdeveloped areas during the different periods.

As shown in Table 6, there were significant (including marginally significant) interactions of area and time on 2 positive SWB indicators and 5 negative SWB indicators.

Table 6. Repeated measures analysis of variance on subjective well-being in the residential lockdown areas.

Indicator	Residential lockdown area, mean (SD)				Area		Time		Area × Time	
	Developed (n=1352)		Underdeveloped (n=438)		$F_{1,1788}$	*P* value	$F_{1,1788}$	*P* value	$F_{1,1788}$	*P* value
	T-before[a]	T-after[b]	T-before	T-after						
Purpose in life	12.19 (1.23)	12.14 (1.30)	11.96 (1.20)	12.08 (1.53)	5.548	.02	0.698	.40	5.057	.03
Autonomy	10.92 (1.81)	10.84 (1.75)	10.57 (1.85)	10.67 (1.79)	8.351	.004	0.099	.75	4.984	.03
Interpersonal sensitivity	9.31 (2.73)	9.55 (2.74)	9.81 (2.97)	9.63 (2.96)	4.342	.04	0.248	.62	9.850	.002
Anxiety	9.48±3.85	9.99±4.16	10.00 (3.95)	10.04±4.42	1.880	.17	9.527	.002	7.273	.007
Hostility	5.44 (2.00)	5.60 (2.08)	5.75 (2.22)	5.66 (2.13)	3.382	.07	0.466	.50	5.202	.02
Paranoid ideation	5.43 (1.42)	5.61 (1.45)	5.57 (1.55)	5.61 (1.47)	1.155	.28	7.822	.005	3.083	.08
Psychoticism	8.63 (2.55)	8.84 (2.83)	9.03 (2.89)	8.97 (2.69)	4.251	.04	1.074	.30	3.584	.06

[a] T-before: the period before the residential lockdown.

[b] T-after: the period after the residential lockdown.

Purpose in Life

The main effect of area was significant ($F_{1,1788}$=5.548, P=.02, η_p^2=0.003). The main effect of time was not significant ($F_{1,1788}$=0.698, P=.40, η_p^2=0.000). The interaction of area and time was significant ($F_{1,1788}$=5.057, P=.03, η_p^2=0.003). After simple effect analysis, our findings were as follows:

Area: In the T-before period, the purpose in life of residents of developed areas with residential lockdowns was significantly higher than that of residents of underdeveloped areas with residential lockdowns ($F_{1,1788}$=11.363, P=.001, η_p^2=0.006); in the T-after period, there was no significant difference between developed and underdeveloped areas with residential lockdowns.

Time: After the implementation of the residential lockdown, there were no significant changes in either developed or underdeveloped areas with residential lockdowns.

Autonomy

The main effect of area was significant ($F_{1,1788}$=8.351, P=.004, η_p^2=0.005). The main effect of time was not significant ($F_{1,1788}$=.099, P=.75, η_p^2=0.000). The interaction of area and time was significant ($F_{1,1788}$=4.984, P=.03, η_p^2=0.003). After simple effect analysis, our findings were as follows:

Area: In the T-before period, the autonomy of residents of developed areas with residential lockdowns was significantly higher than that of residents of under-developed areas with residential lockdowns ($F_{1,1788}$=12.229, P<.001, η_p^2=0.007; in the T-after period, there was no significant difference between developed and underdeveloped areas with residential lockdown.

Time: After the implementation of residential lockdown, there were no significant changes in either developed or underdeveloped areas with residential lockdowns.

Interpersonal Sensitivity

The main effect of area was significant ($F_{1,1788}$=4.342, P=.04, η_p^2=0.002). The main effect of time was not significant ($F_{1,1788}$=0.248, P=.62, η_p^2=0.000). The interaction of area and time was significant ($F_{1,1788}$= 9.850, P=.002, η_p^2=0.005). After simple effect analysis, our findings were as follows:

Area: In the T-before period, the interpersonal sensitivity of residents of developed areas with residential lockdowns was significantly lower than that of residents of underdeveloped areas with residential lockdowns ($F_{1,1788}$=10.643, P=.001, η_p^2=0.006); in the T-after period, there was no significant difference between developed and underdeveloped areas with residential lockdowns.

Time: After the implementation of residential lockdown, the interpersonal sensitivity of residents in developed areas with residential lockdowns increased significantly ($F_{1,1788}$=13.514, P<.001, η_p^2=0.008); meanwhile, there was no significant change in underdeveloped areas with residential lockdowns.

Anxiety

The main effect of area was not significant ($F_{1,1788}$=1.880, P=.17, η_p^2=0.001). The main effect of time was significant ($F_{1,1788}$=9.527, P=.002, η_p^2=0.005). The interaction of area and time was significant ($F_{1,1788}$=7.273, P=.007, η_p^2=0.004). After simple effect analysis, our findings were as follows:

Area: In the T-before period, the anxiety of residents of developed areas with residential lockdowns was significantly lower than that of residents of underdeveloped areas with residential lockdowns ($F_{1,1788}$=5.923, P=.02, η_p^2=0.003); in the T-after period, there was no significant difference between developed and underdeveloped areas with residential lockdowns.

Time: After the implementation of residential lockdown, the anxiety of residents of developed areas with residential lockdowns increased significantly ($F_{1,1788}$=34.173, P<.001, η_p^2=.019); meanwhile, there was no significant change in underdeveloped areas with residential lockdowns.

Hostility

The main effect of area was marginally significant ($F_{1,1788}$=3.382, P=.07, η_p^2=0.002). The main effect of time was not significant ($F_{1,1788}$=0.466, P=.50, η_p^2=0.000). The interaction of area and time was significant ($F_{1,1788}$=5.202, P=.02, η_p^2=0.003). After simple effect analysis, our findings were as follows:

Area: In the T-before period, the hostility of residents of developed areas with residential lockdowns was significantly lower than that of residents of underdeveloped areas with residential lockdowns ($F_{1,1788}$=7.487, P=.006, η_p^2=0.004); in the T-after period, there was no significant difference between developed and underdeveloped areas with residential lockdowns.

Time: After the implementation of residential lockdown, the hostility of residents of developed areas with residential lockdowns increased significantly ($F_{1,1788}$=8.974, P=.003, η_p^2=0.005); meanwhile, there was no significant change in underdeveloped areas with residential lockdowns.

Paranoid Ideation

The main effect of area was not significant ($F_{1,1788}$=1.155, P=.28, η_p^2=0.001). The main effect of time was significant ($F_{1,1788}$=7.822, P=.005, η_p^2=0.004). The interaction of area and time was marginally significant ($F_{1,1788}$=3.083, P=.08, η_p^2=0.002). After simple effect analysis, our findings were as follows:

Area: In both periods, there were no significant differences between developed and underdeveloped areas with residential lockdowns.

Time: After the implementation of residential lockdown, the paranoid ideation of residents of developed areas with residential lockdowns increased significantly ($F_{1,1788}$=21.177, P<.001, η_p^2=0.012); meanwhile, there was no significant change in underdeveloped areas with residential lockdowns.

Psychoticism

The main effect of area was significant ($F_{1,1788}$=4.251, P=.04, η_p^2=.002). The main effect of time was not significant ($F_{1,1788}$=1.074, P=.30, η_p^2=0.001). The interaction of area and time was marginally significant ($F_{1,1788}$=3.584, P=.06, η_p^2=0.002). After simple effect analysis, our findings were as follows:

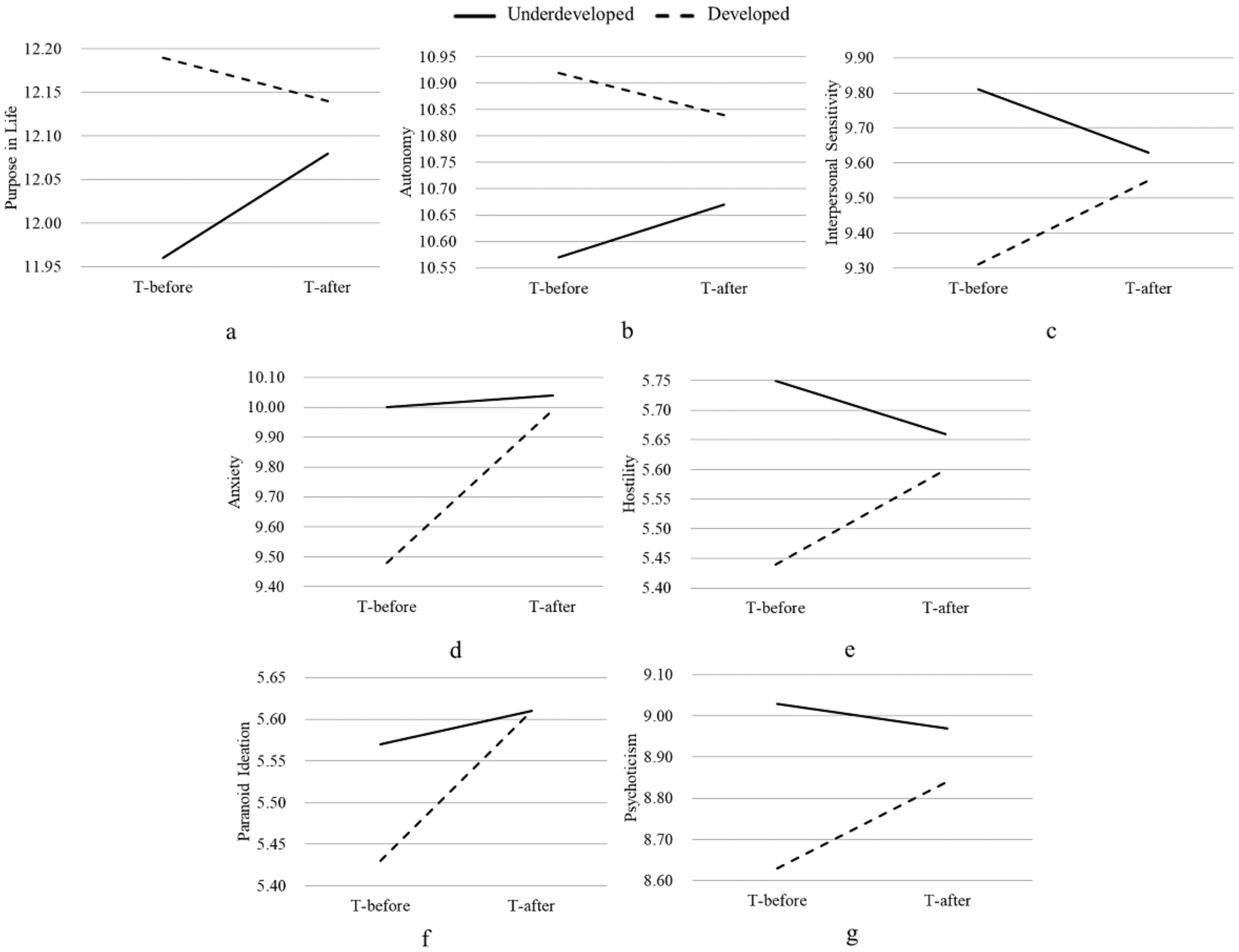
Area: In the T-before period, the psychoticism of residents of developed areas with residential lockdowns was significantly lower than that of residents of underdeveloped areas with residential lockdowns ($F_{1,1788}$=7.822, P=.005, η_p^2=0.004); in the T-after period, there was no significant difference between developed and underdeveloped areas with residential lockdowns.

Time: After the implementation of residential lockdown, the psychoticism of residents of developed areas with residential lockdowns increased significantly ($F_{1,1788}$=8.768, P=.003, η_p^2=0.005); meanwhile, there was no significant change in underdeveloped areas with residential lockdowns.

However, for self-acceptance, environmental mastery, positive relations, personal growth, somatization, obsessive-compulsive, depression, and phobic anxiety, the interactions of area and time were not significant (all *P* values >.10).

The interactions of purpose in life, autonomy, interpersonal sensitivity, anxiety, hostility, paranoid ideation, and psychoticism are presented in Figure 2.

Figure 2. Interactions of area and time on (a) purpose in life, (b) autonomy, (c) interpersonal sensitivity, (d) anxiety, (e) hostility, (f) paranoid ideation, and (g) psychoticism in the residential lockdown areas. T-after: the period after the residential lockdown; T-before: the period before the residential lockdown.



Discussion

Principal Findings

This study aimed to investigate the impact of the residential lockdown on SWB. We obtained the SWB indices of residents inside and outside cities with residential lockdown before and after the enforcement of the residential lockdown policy through OER. In addition, by introducing economic factors, we also made a transverse comparison of SWB in residential lockdown cities divided by economic development level. The results of the RM ANOVAs revealed that the lockdown group scored lower on negative SWB indicators than the nonlockdown group. Moreover, the SWB of residents in underdeveloped areas remained relatively stable after the implementation of the residential lockdown. On the whole, the strict regulations of the residential lockdown and the limitation on residents' travel did have an impact on SWB; however, this impact was positive, and the degree of positivity was greater in underdeveloped areas.

By comparing the residential lockdown cities with corresponding cities without residential lockdowns all over China, the main effect of time from RM ANOVA indicated that as time progressed, the overall SWB of the residents was adversely affected to some degree by the epidemic. This impact was mainly manifested by a general decrease in the scores of positive SWB indicators and an increase in the scores of negative SWB indicators. These results are consistent with previous studies: COVID-19 has caused public pressure and panic owing to the high degree of uncertainty and infectivity of the disease, which has had a deleterious effect on people's mental health [4,14].

However, it is worth mentioning that the simple effect of the interaction in the integral analysis indicated that the implementation of the residential lockdown had a positive influence on the control of residents' negative mental states. Specifically, compared with the nonlockdown group, residents in the lockdown group scored lower on negative SWB indicators. This finding also suggests that stringent residential lockdowns during a pandemic will not only ensure residents' health and life but will also effectively control the prevailing negative mental states of residents. Our findings are in keeping with those of Suryawanshi and More [43], who reported that the positive impact of the lockdown was far greater than the negative impact. Two factors may explain the above results. First, because the residential lockdowns were implemented only after the outbreak of COVID-19, people had a full understanding of the epidemic and were aware that only residential lockdown could effectively stop the spread of the virus by cutting off the transmission route. Therefore, on balance, because the powerful residential lockdown slowed the spread of the virus and curbed the rapid deterioration of the epidemic, it provided greater psychological comfort to the residents compared with the discomfort caused by restriction of personal freedom [43]. Second, the residential lockdown offered family members an opportunity to reevaluate and improve their family relations [43], which could effectively relieve their state of mind.

Furthermore, we used GDP per capita as an indicator of economic development level to classify the cities under residential lockdown. After performing RM ANOVA, we found that the positive impact of the residential lockdown was different in areas with different economic development levels. The SWB of residents in underdeveloped areas with residential lockdown (eg, Huanggang, Huangshi) was found to be relatively stable over time. In contrast, the results showed that during the lockdown in developed areas (eg, Wuhan, Zhongshan), the SWB of the residents generally changed significantly for the worse, although two SWB indicators did not show significant changes (ie, purpose in life and autonomy). The difference in the stability of residents' mental states between the residential lockdown areas at high and low economic development levels can mainly be explained by the following reasons.

Urban and Rural Population Ratio

Economists generally believe that regional economic growth is related to population structure (eg, urban and rural population ratio), among which GDP per capita is positively associated with urban population size [44]. In the residential lockdown areas with higher economic development, the urban population size is larger, and the problems produced by the residential lockdown (eg, unemployment) had a strong impact on urban commuters, which posed great risks to local residents' SWB.

Communication and Transportation

The statistical yearbooks of the residential lockdown cities reported that in the more economically developed areas, key indicators (eg, passenger traffic, number of civil vehicles) would score higher. For example, in 2017, the passenger traffic of Wuhan was 299.5 million, and the number of civil vehicles was 2.75 million [45]. These figures show that residents of more economically developed areas may have had greater transport demand and therefore were more affected by stringent residential lockdown due to the restrictions on their travel.

Floating Population

Economic development is the decisive factor in the formation of population flow. The population usually flows to developed areas with better economic development [46]. During the residential lockdown, the original pattern of life was disrupted, face-to-face association outside home was cut off, and residents mainly obtained social support from family members [47]. However, most members of the floating population make their living without the company of family. Hence, in more economically developed areas, the size of the floating population without family is larger, and the psychological impact of the residential lockdown on residents was more significant.

With the fast rise in the number of confirmed cases, COVID-19 has become a global pandemic. The decision of whether to enforce residential lockdowns is a difficult challenge faced by governments worldwide. The above results indicate that residential lockdowns not only have important medical value in protecting public health and life safety but may also have psychological value in maintaining SWB in China. Furthermore, in developed areas, the urban and floating populations are larger, and the relationship between transportation and life is more important; thus, the SWB of local residents is more likely to be affected by the residential lockdown and fluctuate greatly. These findings provide new references for local governments of areas at different economic development levels to consider the benefit and cost of residential lockdown and to prepare for the possible risks. For example, local governments in developed areas should make more preparations for psychological interventions and associated costs during and following residential lockdown.

To date, Britain, America, and many other countries have adopted lockdowns that are similar to those in China to some extent. Despite cultural differences, the effectiveness of residential lockdown in stopping disease spread is obvious [48]. However, the positive and negative psychological effects of the lockdown are complicated. These effects will also be influenced by the enforcement time, the effects of lockdown, and other factors mentioned above. In addition, in view of the fact that the level of economic development is a common indicator, different regions of the world will also have differences in urban and rural population ratios, floating population, communication and transportation, etc, which may influence the psychological effects caused by residential lockdown. Our findings with regard to the overall effect of the residential lockdown on SWB and the differences in the impact on mental states under different economic development conditions are instructive to analogous policies around the world.

Limitations

Although this study reveals changes in SWB caused by residential lockdown, there are some limitations to be highlighted. First, considering that most of the respondents in our sample are female, there may be some bias in the results. Second, more than 70% of our sample did not provide age information on the internet, which prevented us from evaluating the possible influence of age. In future studies, obtaining more age information and expanding the ranges of age and gender could be helpful. Third, in this study, transverse comparison of SWB in cities divided by economic development level only targeted residential lockdown areas; therefore, in future studies, this comparison can be extended to areas without residential lockdowns. Fourth, in light of the fact that not only economic factors but other factors such as social environment and cultural background may affect SWB, future studies could consider more factors to obtain a more comprehensive understanding of the lockdown effect. Additionally, it should be noted that the patterns we observed may be largely determined by the Chinese culture, and any application of our conclusions to a different culture should be made very carefully.

Conclusion

In this study, to explore the impact of residential lockdown during the COVID-19 pandemic on individual SWB in China, we predicted residents' SWB by OER and compared the SWB of residents of different areas with different economic development levels. The results demonstrated that in China and in the context of the COVID-19 pandemic, the residential lockdown alleviated the negative psychological changes caused by the epidemic to a certain extent. In addition, compared with developed areas, the psychological benefit of the residential lockdown was more obvious in underdeveloped areas. This study sheds light on the psychological effects of strict residential lockdown during pandemics and provides new clues for epidemic prevention in areas with different socioeconomic situations.

Acknowledgments

This study is financially supported by National Natural Science Foundation of China (31700984), Youth Innovation Promotion Association CAS, National Key Research & Development Program of China (2016YFC1307200), and China Social Science Fund (17AZD041).

Authors' Contributions

NZ and TZ conceived and planned the article. NZ is the principal investigator and the guarantor of the paper. YW and NZ carried out the study and drafted the manuscript. DJ and PW collected the data. YW, PW, and SL analyzed the data. NZ, XL, YW, and TZ reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1MB - jmir_v22i12e24775_app1.pdf]

References

1. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
2. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020 May;20(5):553-558. [doi: 10.1016/s1473-3099(20)30144-4]
3. Rubin GJ, Wessely S. The psychological effects of quarantining a city. BMJ 2020 Jan 28;368:m313. [doi: 10.1136/bmj.m313] [Medline: 31992552]
4. Brodeur A, Clark A, Fleche S, Powdthavee N. Assessing the impact of the coronavirus lockdown on unhappiness, loneliness, and boredom using Google Trends. ArXiv Preprint posted online on April 25, 2020.
5. Shojaee M, French C. The Relationship between Mental Health Components and Locus of Control in Youth. PSYCH 2014;05(08):966-978. [doi: 10.4236/psych.2014.58107]
6. Promoting mental health: concepts, emerging evidence, practice: a report of the World Health Organization, Department of Mental Health and Substance Abuse in collaboration with the Victorian Health Promotion Foundation and the University of Melbourne. World Health Organization. 2005. URL: https://apps.who.int/iris/handle/10665/43286 [accessed 2020-12-14]
7. Diener E, Ryan K. Subjective Well-Being: A General Overview. S Afr J Psychol 2009 Dec;39(4):391-406. [doi: 10.1177/008124630903900402]
8. Diener E. Subjective well-being. Psychol Bull 1984 May;95(3):542-575. [Medline: 6399758]
9. Tong Y. Research on general self-efficacy and subjective well-being of low-income college students. Article in Chinese. Chin J Clin Psychol 2003;11(4):294-295. [doi: 10.3969/j.issn.1005-3611.2003.04.018]
10. Xiao J, Shi G. A correlation study of high school students' life events, mental health and subjective well-being. Article in Chinese. J Hebei Normal Univ (Educational Science Edition) 2005;7(2):75-78. [doi: 10.3969/j.issn.1009-413X.2005.02.016]
11. Xing Z. Research of subjective well-being: an empirical exploration of happiness. Article in Chinese. Theory Journal 2002;5:a. [doi: 10.3969/j.issn.1002-3909.2002.05.021]
12. Ma L. Public service performance and citizens' well-being: an empirical study of China's prefecture-level cities. Article in Chinese. Chinese Pubilic Administration 2013;2:a. [doi: 10.3782/j.issn.1006-0863.2013.02.22]
13. Rothstein B, Teorell J. What Is Quality of Government? A Theory of Impartial Government Institutions. Governance 2008 Apr;21(2):165-190. [doi: 10.1111/j.1468-0491.2008.00391.x]

14. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psychiatr 2020;33(2):e100213 [FREE Full text] [doi: 10.1136/gpsych-2020-100213] [Medline: 32215365]
15. Bu H, Liao L. Rural basic-level government's credit and its executive force. Article in Chinese. Journal of Shandong Administration Institute & Shandong Economic Management Personnel Institute 2007;3:9-11. [doi: 10.3969/j.issn.1008-3154.2007.03.003]
16. Ryff CD. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. J Pers Soc Psychol 1989;57(6):1069-1081. [doi: 10.1037/0022-3514.57.6.1069]
17. Xing Z. Research overview of measurement of subjective well-being. Article in Chinese. Psychol Sci 2002;25(3):342-338. [doi: 10.3969/j.issn.1671-6981.2002.03.022]
18. Wan L, Xia L. Relationship of well-being and perfect personality. Article in Chinese. J Southwest China Normal Univ (Humanities and Social Sciences Edition) 2004 Nov;30(6):19-21. [doi: 10.3969/j.issn.1673-9841.2004.06.004]
19. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale--preliminary report. Psychopharmacol Bull 1973 Jan;9(1):13-28. [Medline: 4682398]
20. Holi MM, Sammallahti PR, Aalberg VA. A Finnish validation study of the SCL-90. Acta Psychiatr Scand 1998 Jan;97(1):42-46. [doi: 10.1111/j.1600-0447.1998.tb09961.x] [Medline: 9504702]
21. Holi MM, Marttunen M, Aalberg V. Comparison of the GHQ-36, the GHQ-12 and the SCL-90 as psychiatric screening instruments in the Finnish population. Nord J Psychiatry 2003;57(3):233-238. [doi: 10.1080/08039480310001418] [Medline: 12775300]
22. Buchanan T, Smith JL. Using the Internet for psychological research: personality testing on the World Wide Web. Br J Psychol 1999 Feb;90 (Pt 1):125-144. [doi: 10.1348/000712699161189] [Medline: 10085550]
23. Carlbring P, Brunt S, Bohman S, Austin D, Richards J, Öst L, et al. Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. Comput Hum Behav 2007 May;23(3):1421-1434. [doi: 10.1016/j.chb.2005.05.002]
24. Bishop J, Martin A, Costanza S, Lane R. Suicide Signs on the Rorschach. J Contemp Psychother 2000;30(3):289-305. [doi: 10.1023/A:1004146800590]
25. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, et al. Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence. J Med Internet Res 2020 Aug 18;22(8):e22590 [FREE Full text] [doi: 10.2196/22590] [Medline: 32750001]
26. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci USA 2013 Apr 09;110(15):5802-5805 [FREE Full text] [doi: 10.1073/pnas.1218772110] [Medline: 23479631]
27. Qiu L, Lin H, Ramsay J, Yang F. You are what you tweet: Personality expression and perception on Twitter. J Res Pers 2012 Dec;46(6):710-718. [doi: 10.1016/j.jrp.2012.08.008]
28. Hao B, Li A, Bai S, Zhu T. Predicting psychological features based on web behavioral data: Mental health status and subjective well-being. Chin Sci Bull 2015 Apr 9;60(11):994-1001. [doi: 10.1360/n972014-00763]
29. Liu M, Xue J, Zhao N, Wang X, Jiao D, Zhu T. Using Social Media to Explore the Consequences of Domestic Violence on Mental Health. J Interpers Violence 2018 Feb 01:886260518757756. [doi: 10.1177/0886260518757756] [Medline: 29441804]
30. Li L, Li A, Hao B, Guan Z, Zhu T. Predicting active users' personality based on micro-blogging behaviors. PLoS One 2014;9(1):e84997 [FREE Full text] [doi: 10.1371/journal.pone.0084997] [Medline: 24465462]
31. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. Am Psychol 2015 Sep;70(6):543-556. [doi: 10.1037/a0039210] [Medline: 26348336]
32. Easterlin RA. Will raising the incomes of all increase the happiness of all? J Econ Behav Organ 1995 Jun;27(1):35-47. [doi: 10.1016/0167-2681(95)00003-b]
33. Easterlin RA. Income and Happiness: Towards a Unified Theory. Econ J 2001;111(473):465-484. [doi: 10.1111/1468-0297.00646]
34. Guan H. The impact of income on happiness: absolute and relative measures. Article in Chinese. Nankai Econ Stud 2010(5):56. [doi: 10.3969/j.issn.1001-4691.2010.05.005]
35. Zhu J, Yang X. An empirical study on income and well-being in China's transitional period. Article in Chinese. Stat Res 2009;26(4):7-12. [doi: 10.3969/j.issn.1002-4565.2009.04.002]
36. Clark A, Senik C. Will GDP growth increase happiness in developing countries? IZA Discussion Paper No 5595. SSRN Journal Preprint posted online on March 28, 2011.
37. Inglehart R, Foa R, Peterson C, Welzel C. Development, Freedom, and Rising Happiness: A Global Perspective (1981-2007). Perspect Psychol Sci 2008 Jul;3(4):264-285. [doi: 10.1111/j.1745-6924.2008.00078.x] [Medline: 26158947]
38. Hao B, Li L, Li A, Zhu T. Predicting Mental Health Status on Social Media. In: Rau PLP, editor. Cross-Cultural Design. Cultural Differences in Everyday Life. Cham, Switzerland: Springer; 2013:101-110.

39. Gao R, Hao B, Li H, Gao Y, Zhu T. Developing simplified Chinese psychological linguistic analysis dictionary for microblog. In: Imamura K, Usui S, Shirao T, Kasamatsu T, Schwabe L, Zhong N, editors. Brain and Health Informatics. BHI 2013. Lecture Notes in Computer Science, vol 8211. Cham, Switzerland: Springer; 2013:359-368.
40. Liu T, Che W, Li Z. The linguistic platform technology. Article in Chinese. Journal of Chinese Information Processing 2011;6:53-62. [doi: 10.3969/j.issn.1003-0077.2011.06.008]
41. Zhao N, Jiao D, Bai S, Zhu T. Evaluating the Validity of Simplified Chinese Version of LIWC in Detecting Psychological Expressions in Short Texts on Social Network Services. PLoS One 2016;11(6):e0157947 [FREE Full text] [doi: 10.1371/journal.pone.0157947] [Medline: 27322382]
42. Jin L. The relationship of subjective well-being and psychological well-being. Dissertation in Chinese. Hebei Normal University 2007. [doi: 10.7666/d.y1099608]
43. Suryawanshi R, More V. A Study of Effect of Corona Virus Covid-19 and Lock Down on Human Psychology of Pune City Region. Studies in Indian Place Names 2020 Mar;40(70):984-994.
44. Liu R, Zhang A, Ding X, Gao X. Research on the economic growth effect of the county population structure in Qinghai Province. Article in Chinese. Statistics and Decision-Making 2015;10:107-110. [doi: 10.13546/j.cnki.tjyjc.2015.10.031]
45. Meng WK, Chen XQ, editors. Wuhan Statistical Yearbook. Beijing, China: China Statistics Press; 2018.
46. Wang X. In-depth exploration of the relationship between floating population and regional economic and social development ? a review of Research on floating population and economic and social development: taking Jilin province as an example. Article in Chinese. Population Journal 2016;38(5):112. [doi: 10.3969/j.issn.1004-129X.2016.05.012]
47. Kessler RC, Price RH, Wortman CB. Social factors in psychopathology: stress, social support, and coping processes. Annu Rev Psychol 1985;36:531-572. [doi: 10.1146/annurev.ps.36.020185.002531] [Medline: 3883893]
48. Khanna R, Cicinelli M, Gilbert S, Honavar S, Murthy GV. COVID-19 pandemic: Lessons learned and future directions. Indian J Ophthalmol 2020;68(5):703. [doi: 10.4103/ijo.ijo_843_20]

Abbreviations

GDP: gross domestic product
OER: online ecological recognition
PWB: Psychological Well-Being scale
RM ANOVA: repeated measures analysis of variance
SCL-90: Symptom Checklist 90
SCLIWC: Simplified Chinese Language Inquiry and Word Count
SWB: subjective well-being
WHO: World Health Organization

Edited by G Eysenbach; submitted 07.10.20; peer-reviewed by A Li, J Xue; comments to author 30.10.20; revised version received 19.11.20; accepted 05.12.20; published 17.12.20.

Please cite as:
Wang Y, Wu P, Liu X, Li S, Zhu T, Zhao N
Subjective Well-Being of Chinese Sina Weibo Users in Residential Lockdown During the COVID-19 Pandemic: Machine Learning Analysis
J Med Internet Res 2020;22(12):e24775
URL: http://www.jmir.org/2020/12/e24775/
doi:10.2196/24775
PMID:33290247

Original Paper

Smartphone-Based Virtual Agents to Help Individuals With Sleep Concerns During COVID-19 Confinement: Feasibility Study

Pierre Philip[1,2,3], Prof Dr; Lucile Dupuy[1,3], PhD; Charles M Morin[4,5], PhD; Etienne de Sevin[1,3], PhD; Stéphanie Bioulac[1,2,3], MD, PhD; Jacques Taillard[1,3], PhD; Fuschia Serre[1,3], PhD; Marc Auriacombe[1,3], Prof Dr; Jean-Arthur Micoulaud-Franchi[1,2,3], MD, PhD

[1]USR 3413 SANPSY, University of Bordeaux, Bordeaux, France

[2]Service de Médecine du Sommeil, University Hospital of Bordeaux, Bordeaux, France

[3]SANPSY, USR 3413, Centre Nationale de la Recherche Scientifique, Bordeaux, France

[4]Ecole de psychologie, Université Laval, Quebec, QC, Canada

[5]Centre d'étude des troubles du sommeil, Centre de recherche CERVO, Institut universitaire en santé mentale de Québec, Québec, QC, Canada

Corresponding Author:
Lucile Dupuy, PhD
USR 3413 SANPSY
University of Bordeaux
Bordeaux
France
Phone: 33 557571100
Email: lucile.dupuy@u-bordeaux.fr

Abstract

Background: The COVID-19 crisis and consequent confinement restrictions have caused significant psychosocial stress and reports of sleep complaints, which require early management, have increased during recent months. To help individuals concerned about their sleep, we developed a smartphone-based app called KANOPEE that allows users to interact with a virtual agent dedicated to autonomous screening and delivering digital behavioral interventions.

Objective: Our objective was to assess the feasibility of this app, in terms of inclusion rate, follow-up rate, perceived trust and acceptance of the virtual agent, and effects of the intervention program, in the context of COVID-19 confinement in France.

Methods: The virtual agent is an artificial intelligence program using decision tree architecture and interacting through natural body motion and natural voice. A total of 2069 users aged 18 years and above downloaded the free app during the study period (April 22 to May 5, 2020). These users first completed a screening interview based on the Insomnia Severity Index (ISI) conducted by the virtual agent. If the users were positive for insomnia complaints (ISI score >14), they were eligible to join the 2-stage intervention program: (1) complete an electronic sleep diary for 1 week and (2) follow personalized sleep recommendations for 10 days. We collected and analyzed the following measures: sociodemographic information, ISI scores and sleep/wake schedules, and acceptance and trust of the agent.

Results: Approximately 76% (1574/2069) of the app users completed the screening interview with the virtual agent. The virtual agent was well accepted by 27.4% (431/1574) of the users who answered the acceptance and trust questionnaires on its usability, satisfaction, benevolence, and credibility. Of the 773 screened users who reported sleep complaints (ISI score >14), 166 (21.5%) followed Step 1 of the intervention, and only 47 of those (28.3%) followed Step 2. Users who completed Step 1 found that their insomnia complaints (baseline mean ISI score 18.56, mean ISI score after Step 1 15.99; $P<.001$) and nocturnal sleep quality improved significantly after 1 week. Users who completed Step 2 also showed an improvement compared to the initial measures (baseline mean ISI score 18.87, mean ISI score after Step 2 14.68; $P<.001$). Users that were most severely affected (ISI score >21) did not respond to either intervention.

Conclusions: These preliminary results suggest that the KANOPEE app is a promising solution to screen populations for sleep complaints and that it provides acceptable and practical behavioral advice for individuals reporting moderately severe insomnia.

(J Med Internet Res 2020;22(12):e24268) doi:10.2196/24268

KEYWORDS

COVID-19; virtual agent; sleep disorders; technology acceptance; agent; sleep; smartphone; mobile phone; eHealth; feasibility; stress; app; intervention; behavior

Introduction

The current COVID-19 crisis has led to massive public health interventions, resulting in the confinement of almost the entire human population worldwide [1]. However, recent studies suggest that confinement may induce several negative psychological effects, including post-traumatic stress symptoms, anxiety, depression, anger, and insomnia [2,3]. Notably, Voitsidis et al [4] showed that in a study comprising 2363 Greek subjects, almost 38% of them reported insomnia complaints during the confinement due to COVID-19, and these complaints were associated with a higher rate of depression. The use of tobacco and alcohol in association with depression and stress symptoms has also increased during the COVID-19 pandemic [5].

These findings confirm that the COVID-19 crisis has caused major psychosocial stress and that prolonged confinement is potentially an aggravating factor for sleep complaints and insomnia. Therefore, given the large number of individuals affected and the limited number of health care professionals available during the crisis, there is a need for innovative solutions to track and help individuals at risk of psychosocial stress.

Digital technologies play a significant role in the context of the ongoing pandemic and overwhelmed health care services. As suggested by many researchers [6,7] and governmental authorities [1], technologies such as social media, smartphone apps, telehealth, and big data have great potential to disseminate information as well as screen and remotely monitor the general population, including patients with COVID-19. Several apps have been deployed in recent times to assist with COVID-19 management (eg, STOP COVID in France [8]). However, only a few apps and technologies specifically address psychosocial stress induced by the COVID-19 crisis and confinement [9-11]. To our knowledge, no study has focused on insomnia complaints, despite the evidence that digital behavioral therapies are effective to treat insomnia [12-14].

Soon after confinement measures were effective in France on March 17, 2020, we launched the first social media campaign in affiliation with our hospital and university through major national radios and newspapers. This campaign focused on the risk of insomnia and the measures to evaluate and correct inappropriate sleep hygiene practices among people during the COVID-19 confinement. In addition to social media campaigns, we developed a free smartphone app to help individuals with sleep concerns in the context of the COVID-19 pandemic. Named KANOPEE, the program is based on our previous research on embodied conversational agents (ECAs), also called virtual agents, which may be defined as animated characters that can engage in face-to-face dialogue through verbal and nonverbal behaviors. We previously demonstrated that ECAs can deliver a clinical interview to diagnose not only sleep complaints but also addiction and depression in an autonomous, reliable, valid, and acceptable way [15-19], by fostering empathy and facilitating disclosure of negatively connoted topics. In addition, based on the existing tools and knowledge on digital therapies for insomnia [14], we developed a digital sleep diary to automatically quantify the user's daily sleep patterns and sleep duration, and to establish personalized sleep interventions guided by the data collected through the app.

We hypothesized that a virtual agent made available via a smartphone app would be efficient and acceptable not only in providing autonomous screening for insomnia complaints but also in establishing digital behavioral interventions to help the population during the COVID-19 crisis. Therefore, to test our hypothesis, we launched a proof-of-concept study during the COVID-19 confinement.

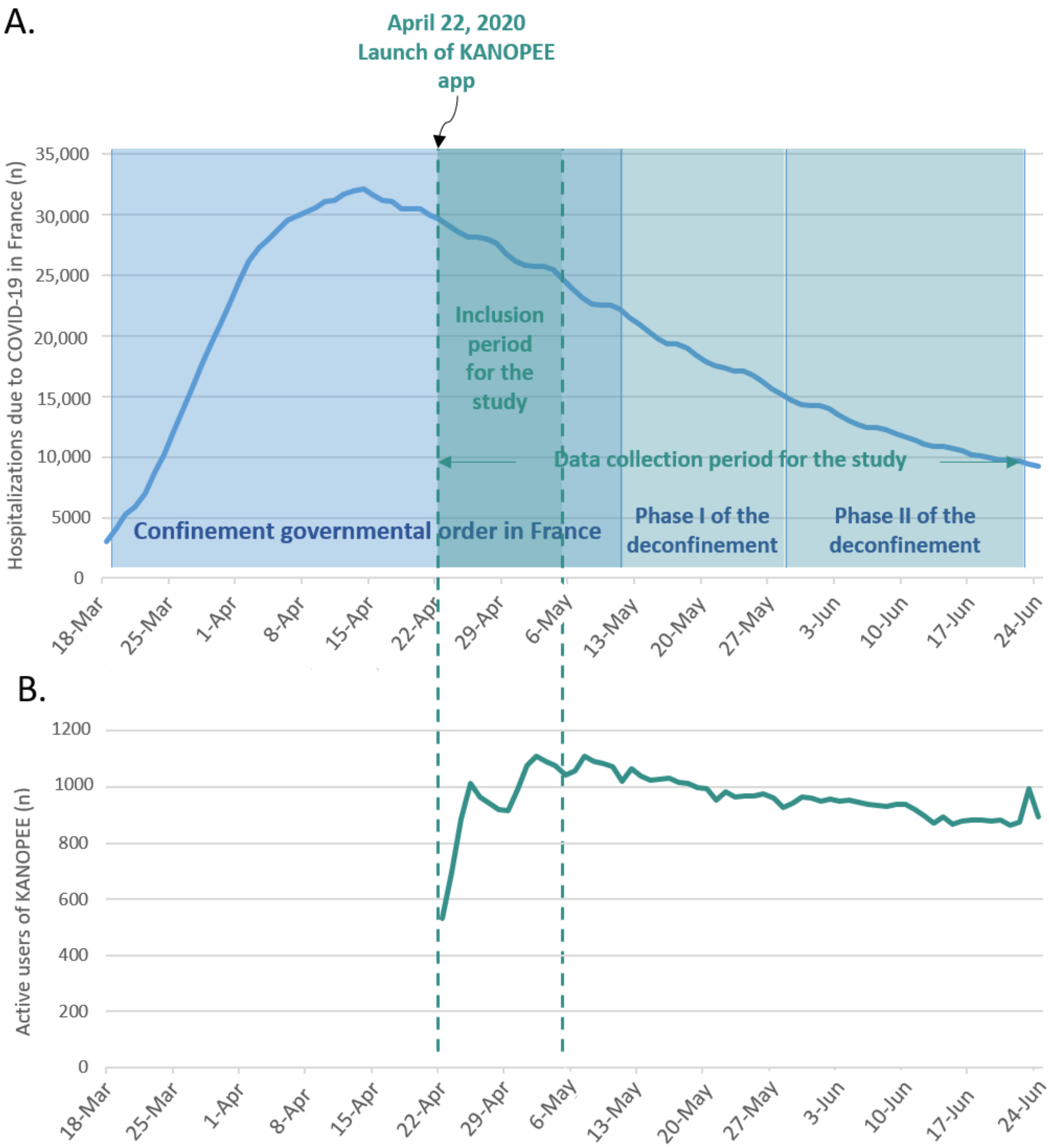
Methods

Description and Implementation of KANOPEE

KANOPEE was implemented using the same architecture as our previous ECAs [15-19], with C# programming language in Unity 3D software (Unity Technologies). The virtual agent, named Louise, interacts through natural voice and body motion, as recorded by a professional artist by using motion capture technologies (Optitrack suit for 3D animations of Louise's body and GoPro for 3D animations of Louise's face and her voice). The interaction scenario is predefined, using decision trees to adapt to the user's answers.

The app was made freely available on Google Play Store on April 22, 2020 (see Figure 1). After its launch, we organized a second media campaign presenting KANOPEE and showing how the app could help the French population to self-evaluate their sleep quality and provide practical solutions to manage insomnia. We analyzed the number of active app users across the confinement and deconfinement stages in France as well as hospitalizations due to COVID-19 during this time (Figure 1, [20])

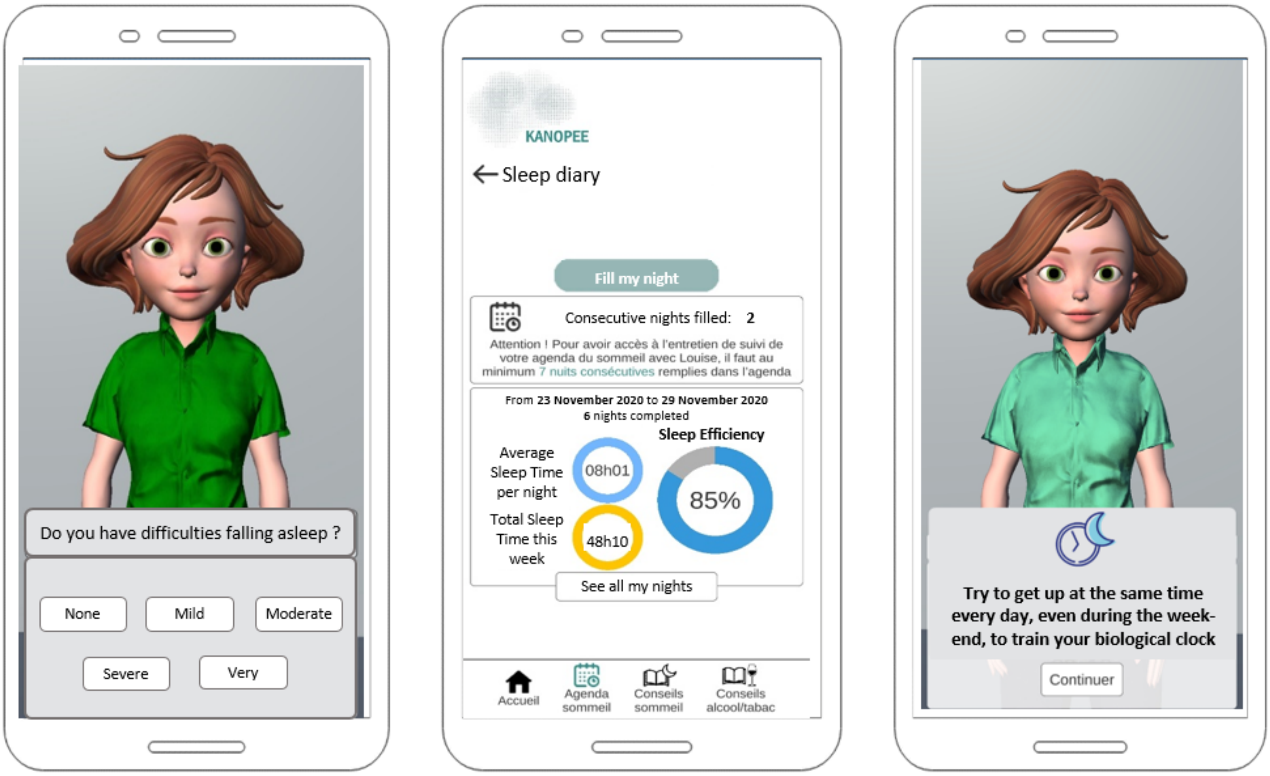
Figure 1. Conception and use of KANOPEE app. Chronological evolution of (A) the number of hospitalizations due to COVID-19 in France [20] and (B) the number of active users after the launch of the app across the confinement and deconfinement stages in France (based on data sourced from the Statistics page of the Google Play Store developer console for the app).



User Interaction

The interaction scenario comprises of the following steps: first, during the screening interview (Interview 1), Louise introduces herself and administers the Insomnia Severity Index (ISI) [21] (see Figure 2, leftmost screen capture). Then, depending on the score, users are either provided with simple sleep hygiene recommendations (eg, follow usual wake up time, get exposed to sunlight in the morning, sleep in a quiet and dark room) if ISI score ≤14, or they are considered eligible to enter the intervention program if ISI score >14 [22]. Initially, users enter the "first step" of the intervention program, wherein they receive instructions to complete a sleep diary for 1 week in order to have a better understanding of their sleep patterns and to collect data about sleep indicators. Every morning, after filling in their sleep schedule, these users receive visual feedback on their sleep (eg, time spent in bed, total sleep time, and sleep efficiency; see Figure 2, middle screen capture). After completing the sleep diary for 7 days, they receive a follow-up interview request with Louise (Interview 2) wherein they learn about their sleep indicators from the previous week. At this point, they receive an ISI score for the second time. Next, the users can enter the "second step," during which they are provided with personalized sleep recommendations for 10 days based on the sleep diary data and their answers on the ISI (see Figure 2, rightmost screen capture). Details of the conditions for providing personalized recommendations are listed in Textbox S1 (Multimedia Appendix 1). Thereafter, they can access another follow-up interview (Interview 3) and take the ISI a third time. Depending on the third ISI score, users can continue to use the app autonomously or, if their sleep complaints persisted (ie, ISI score >21), be referred to a sleep specialist in our hospital.

Figure 2. Illustrative references of KANOPEE app interfaces. From left to right, screenshot of (1) virtual agent Louise conducting an interview based on the Insomnia Severity Index, (2) sleep diary and visual feedbacks regarding user's sleep patterns, and (3) a personalized sleep recommendation given by Louise during Interview 3.



Do you have difficulties falling asleep ?

Consecutive nights filled: 2

Attention ! Pour avoir accès à l'entretien de suivi de votre agenda du sommeil avec Louise, il faut au minimum 7 nuits consécutives remplies dans l'agenda

From 23 November 2020 to 29 November 2020
6 nights completed

Average Sleep Time per night: 08h01

Total Sleep Time this week: 48h10

Sleep Efficiency: 85%

Try to get up at the same time every day, even during the week-end, to train your biological clock

Throughout the process, all procedures and tools (ie, questionnaires and sleep diary) are introduced by the virtual agent in order to facilitate understanding among the users and increase their engagement. A demo video of the user interaction by Louise has been hosted on YouTube [23].

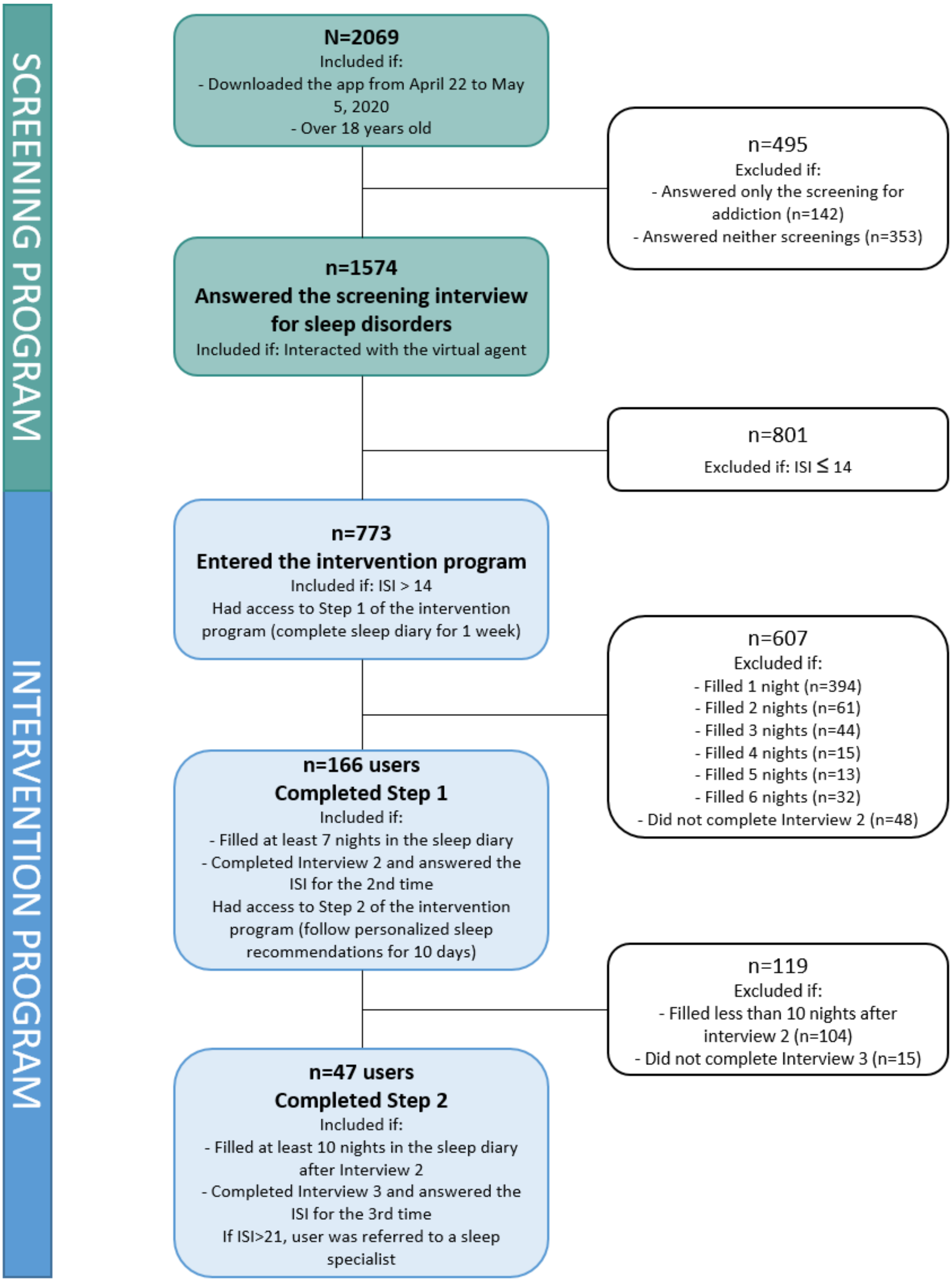
Sociodemographics and Clinical Characteristics of Users

For the purpose of this study, users were selected for the analysis if they met the following inclusion criteria: (1) aged 18 years old and above and (2) had downloaded KANOPEE app before May 5, 2020, such that they had access to the 1-week intervention before the end of confinement period in France (ie, May 11, 2020). Their use of the app was recorded from April 22 until May 26.

After getting approval by the University and Hospital scientific committees, we obtained authorizations to be registered on the University Hospital register for General Data Protection Regulation (GDPR) approval by the French authorities—Commission nationale de l'informatique et des libertés (CNIL). Informed consent was obtained from all users downloading the app according to the GDPR and CNIL regulations.

Subgroups of users were then selected for more detailed analyses (see Figure 3). Specifically, users who answered the screening interview for sleep disorders were further examined to determine their eligibility for the next step of daily sleep monitoring, and users who reported insomnia complaints (ie, ISI score >14) entered the intervention program and were included in the analyses of outcomes and feasibility.

Figure 3. Flowchart of users included in the different steps of analyses.



Measures

Clinical Measures

The ISI [21] is a 7-item self-report questionnaire that provides a global measure of perceived insomnia severity, with scores ranging from 0 to 28 (0-7: no clinical insomnia, 8-14: subthreshold insomnia, 15-21: insomnia of moderate severity, and 22-28: severe insomnia). The ISI has been previously validated and proven sensitive to changes in insomnia severity with treatment [22]. The ISI was used as a screening tool for assessing insomnia severity and as the primary outcome measure of treatment efficacy after the intervention.

Users who had an ISI score >14 were asked to complete a daily sleep diary in the app [24,25] throughout the intervention program. The following dependent variables were derived from users' daily sleep diary entries: sleep onset latency (SOL, ie, how many minutes it takes to fall asleep, starting from the moment of intention to fall asleep), number of awakenings

(NWAK), terminal wakefulness (TWAK, ie, the amount of awake time between the final wakefulness awakening and the time of getting out of bed), wake after initial sleep onset (WASO, ie, total amount of time awake during the night, excluding SOL and TWAK), total time spent in bed (TIB, ie, time starting from the moment of intention to fall asleep and concluding with the final arising), total sleep time (TST, ie, time actually spent sleeping) that was calculated from other self-reported variables (TIB−SOL− WASO−TWAK), and sleep efficiency (ie, percentage of time in bed spent asleep) that was calculated from other self-reported variables (TST/TIB × 100).

Addictive behaviors (ie, alcohol and cigarette consumption) of users were evaluated through a clinical interview based on the CAGE (Cut down, Annoyed, Guilty, and Eye-opener) [26] and CDS-5 (Cigarette Dependence Scale, 5-item) questionnaires [27].

Acceptance and Trust Questionnaires

After the interviews with the virtual agent, users could complete 2 assessments on the app. The first assessment was the French version of the Acceptability E-scale (AES) [28,29] to measure acceptance of the KANOPEE app based on 2 subscores: usability (ie, the perceived ease of using the system or app) and satisfaction (ie, the perceived enjoyment of the use and usefulness of the system or app). The second assessment was the ECA trust questionnaire (ETQ) [18] that measures a user's trust in a virtual agent based on 2 subdimensions: perceived credibility (ie, perception that the agent has the ability and the expertise to conduct a medical intervention) and benevolence (ie, perception that the agent is well-intentioned and will accurately take the user's interests into account). Familiarity with technologies was also evaluated by a single question: “Are you familiar with computer technologies?” with the following 3 choices: “No,” “Moderately,” and “Yes,” which were scored as 0, 1, and 2, respectively.

Statistical Analyses

Quantitative variables were expressed as means and SD, and qualitative variables were expressed as percentages. To compare 2 groups of users (eg, subclinical insomnia vs with moderate-to-severe insomnia), we performed 2-tailed Student *t* tests for continuous variables (eg, age, CDS score, and cigarettes smoked), and χ^2 tests for categorical variables (eg, gender, educational level, health care professional, and users adhering to the confinement). The data collected during the program were described using mean and SD values, and evolution of the measures over time was analyzed using repeated *t* tests. Acceptance (ie, usability and satisfaction subscores of the AES) and trust (ie, credibility and benevolence subscores of the ETQ) data were expressed using distributions and percentages. To investigate factors associated with acceptance and trust, we conducted univariate analyses with Pearson correlation analyses between 2 continuous variables (age, insomnia severity, and familiarity with technologies) and performed mean comparisons (*t* test or analysis of variance) to analyze the variation in AES and ETQ scores regarding categorical variables (gender and educational level). All analyses were performed using SPSS software (version 26; IBM Corp).

Results

Characteristics of App Users

A total of 2069 users aged ≥18 years downloaded KANOPEE app and answered sociodemographic information (Table 1). Of these, 76% (1574/2069) of the users answered the screening interview for insomnia disorders. Most of these users were between 31 and 50 years old (mean 43.52, SD 13.94) and had a university degree. Furthermore, most users were in confinement due to the COVID-19 lockdown, and 5.6% (89/1574) were front-line health care professionals involved in the fight against COVID-19. Approximately half of the users (773/2069, 49.1%) who answered the screening interview for sleep disorders obtained an ISI score over the clinical threshold for insomnia (ie, ISI score >14). Compared to the users who answered the screening interview for sleep, those who chose not to were significantly older (P=.016) and predominantly male users (P=.001). Other factors remained nonsignificant between the 2 groups (see Table 1).

Table 1. Characteristics of KANOPEE users depending on their use of the app.

Characteristics	Values		Group comparison		*P* value
	Answered screening interview for sleep (n=1574)	Did not answer screening interview for sleep (n=495)	*t* test (*df*)	Chi-square (*df*)	
Age, mean (SD)	43.11 (13.8)	44.83 (14.4)	*2.40 (2067)* [a]	N/A[b]	*.02*
Age in years , n (%)					
18-30	338 (21.5)	99 (20)			
31-50	763 (48.5)	220 (44.4)			
51-65	366 (23.3)	131 (26.5)			
>65	107 (6.8)	45 (9.1)			
Gender, n (%)			N/A	*10.71 (1)*	*.001*
Female	1055 (67)	292 (59)			
Male	519 (33)	203 (41)			
Educational level, n (%)			N/A	5.69 (*2*)	.13
Middle school	318 (20.2)	83 (16.8)			
High school	309 (19.6)	118 (23.8)			
University degree	947 (60.2)	294 (59.4)			
Health care professionals, n (%)	89 (5.6)	28 (5.6)	N/A	0.00 (*1*)	.99
Confined due to COVID-19 lockdown, n (%)	1200 (76.2)	371 (74.9)	N/A	0.343 (*1*)	.50

[a] Italicized values indicate they are statistically significant.

[b] N/A: not applicable.

Users who answered the screening interview for sleep disorders (n=1574) were divided into 2 subgroups based on their performance on the ISI: scores ≤14 considered “with subclinical insomnia” and scores >14 considered “with moderate-to-severe insomnia” (Table 2).

Compared to users with subclinical insomnia, users with moderate-to-severe insomnia were younger ($t_{1576}=-3.03$; *P*=.002), more educated ($\chi^2_2=12.14$; *P*=.007), and more likely to be female ($\chi^2_1=31.91$; *P*<.001). Interestingly, more users in confinement were found in the moderate-to-severe insomnia group ($\chi^2_1=8.86$; *P*=.003), but we did not find evidence of a higher prevalence of insomnia among health care professionals. Users with moderate-to-severe insomnia smoked more cigarettes ($t_{734}=-4.03$; *P*<.001) and obtained a higher score on the screening questionnaire for addiction to cigarettes ($t_{734}=3.41$; *P*=.001) than those in the other groups.

Table 2. Characteristics of users depending on their sleep complaints.

Characteristics	Values		Group comparison		*P* value
	ISI[a] score ≤14 (individuals with subclinical insomnia, n=801)	ISI score >14 (individuals with moderate to severe insomnia, n=773)	*t* test (*df*)	Chi-square (*df*)	
Age, mean (SD)	44.1 (14.3)	42.0 (13.1)	*−3.03 (1576)*[b]	N/A[c]	*.002*
Age in years, n (%)					
18-30	166 (20.7)	172 (22.3)			
31-50	372 (46.3)	391 (50.6)			
51-65	194 (24.2)	172 (22.3)			
>65	69 (8.7)	38 (4.9)			
Gender, n (%)			N/A	*31.91 (1)*	*.001*
Female	484 (60.4)	571 (73.9)			
Male	317 (39.6)	202 (26.1)			
Educational level, n (%)			N/A	*12.14 (2)*	*.007*
Middle school	150 (18.6)	168 (21.7)			
High school	139 (17.5)	170 (22.0)			
University degree	512 (63.9)	435 (56.2)			
Health care professionals, n (%)	53 (6.6)	36 (4.7)	N/A	2.75 (*1*)	.097
Confined due to COVID-19 lockdown, n (%)	586 (73.1)	614 (79.4)	N/A	*8.86 (1)*	*.003*
Clinical insomnia, ISI score, mean (SD)	10.02 (3.42)	18.2 (2.74)	*52.12 (1572)*	N/A	*.001*
Clinical insomnia subtype, n (%)					
No clinically significant insomnia (ISI score ≤7)	180 (22.5)	0 (0)			
Subthreshold (ISI score 8-14)	621 (77.5)	0 (0)			
Moderate (ISI score 15-21)	0 (0)	672 (86.9)			
Severe (ISI score ≥21)	0 (0)	101 (13.1)			
CDS-5[d], mean (SD)	4.53 (6.93)	6.45 (8.30)	*3.41 (734)*	N/A	*.001*
Daily number of cigarettes, mean (SD)	3.09 (6.40)	5.33 (8.68)	*−4.03 (734)*	N/A	*.001*
CAGE[e], mean (SD)	0.65 (0.97)	0.78 (1.12)	1.66 (*734*)	N/A	.097
Daily number of drinks, mean (SD)	1.34 (2.09)	1.66 (3.18)	−1.62 (*734*)	N/A	.11

[a] ISI: Insomnia Severity Index.

[b] Italicized values indicate they are statistically significant.

[c] N/A: not applicable.

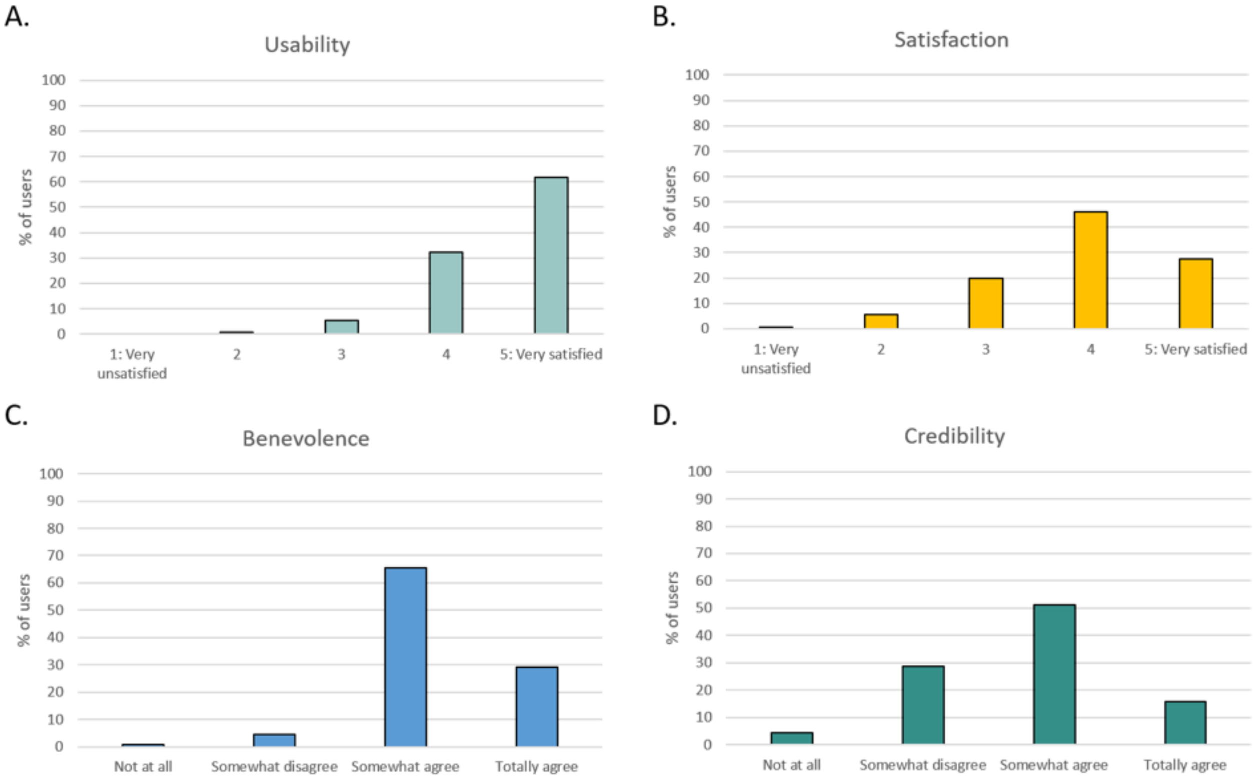
[d] CDS-5: Cigarette Dependence Scale-5 items.

[e] CAGE: Cut-down, Annoyed, Guilty, Eye-opener questionnaire.

Trust and Acceptance of Virtual Agents

In all, 431 of 1574 (27.4%) users answered the acceptance and trust questionnaires (Figure 4). Acceptance of the overall system (AES score) was rated very positively, with 61.7% (266/431) of users being “very satisfied” with the usability of the system, and 91.6% (395/431) of users rating the virtual agent more than 3 out of 5 for satisfaction. Regarding trust (ETQ score), Louise was perceived as trustworthy to perform medical interviews. Indeed, 94.6% (408/431) of users “somewhat agreed” or “totally agreed” that she was benevolent, and 67.05% (289/431) of users had a positive attitude towards her credibility (ie, rating of more than 1 out of 3).

Figure 4. Distribution of usability, satisfaction, benevolence, and credibility perception of the virtual companion for sleep disorders (Louise). (A) Percentage of users' rating for usability dimension (AES subscore), (B) percentage of users' rating for satisfaction dimension (AES subscore), (C) percentage of users' rating for benevolence dimension (ETQ subscore), and (D) percentage of users' rating for credibility dimension (ETQ subscore).

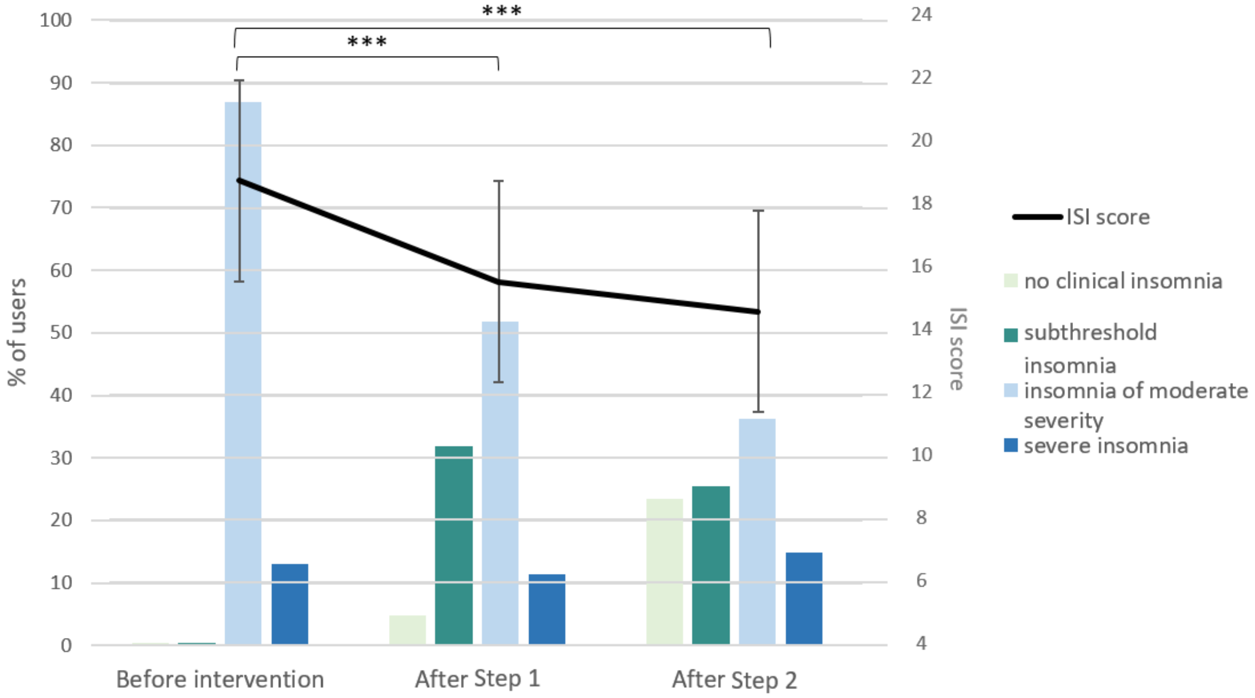


We found a negative correlation between age and credibility subscore on the ETQ (r=−.102; *P*=.034), suggesting that older individuals found Louise less credible than the younger ones. Age was not correlated with other dimensions of trust and acceptance. Similarly, gender and educational level of the users were not correlated with their attitude towards Louise. Regarding insomnia severity, there was a positive relationship between the severity and credibility of Louise (r=.125; *P*=.009), indicating that users with more severe insomnia complaints found her more credible. Lastly, we found significant correlations between users' familiarity with technologies and their attitudes towards Louise: users more familiar with technologies found her more usable (r=.109; *P*=.024), more satisfactory (r=.128; *P*=.008), and more benevolent (r=.117; *P*=.015).

Evolution of ISI Score and Nocturnal Sleep Indicators During the Intervention Program

Among the 166 users who completed Step 1 of the intervention (ie, fill a sleep diary for 1 week and answer the ISI for the second time), the total ISI score (Figure 5) decreased compared to the baseline (baseline mean ISI score 18.56, mean ISI score after Step 1 15.99; t_{165}=7.88; *P*<.001), with 36.7% (61/166) of users obtaining an ISI score below a clinically significant level (ie, ≤14) either corresponding to "no insomnia" (8/166, 4.8%) or to "subthreshold insomnia" (53/166, 31.9%). For the 47 users who completed Step 2 of the intervention, their ISI scores continued to decrease but did not reach a significant threshold (mean ISI score after Step 1 15.64, mean ISI score after Step 2 14.68; t_{46}=1.42; *P*=.162). However, compared to the initial measure, a significant decrease was observed (baseline mean ISI score 18.87, mean ISI score after Step 2 14.68; t_{46}=4.85; *P*<.001). Moreover, the proportion of users reporting low insomnia complaints increased, with a total of 48.9% (23/47) of users below a clinically significant level. Of note, 14.9% (7/47) of users still reported "severe insomnia" after Step 2, so they were referred to a sleep specialist.

Figure 5. Distribution of users depending on the severity of their insomnia complaints (based on ISI scores) along the intervention program: Step 1: completed sleep diary; Step 2: followed personalized sleep recommendations. ISI: Insomnia Severity Index. Significance: *$P<.05$, **$P<.001$.



Regarding nocturnal sleep patterns, we computed the mean scores of the first 2 nights filled in the sleep diary and the last 2 nights before receiving step 2 intervention in order to evaluate the evolution of sleep indicators during completion of Step 1 of the intervention program among the 166 users who completed the Step 1 (ie, filling in the sleep diary). Analyses of mean and SD values (see Table 3) suggest a reduction in TIB, SOL, and TWAK, and an increased sleep efficiency among this subgroup.

Table 3. Nocturnal sleep indicators of users completing Step 1 (n=166).

Sleep indicator	Value, mean (SD)		*t* test (*df*)	*P* value
	First 2 nights (1 & 2)	Last 2 nights (6 & 7)		
Time in bed, hh:mm:ss	08:56:45 (01:34:22)	08:40:27 (01:29:27)	*2.17 (165)*[a]	*.03*
Total sleep time, hh:mm:ss	06:04:32 (01:56:47)	06:21:30 (01:37:33)	−1.88 (165)	.06
Sleep efficiency, %	67.60 (20.25)	73.82 (17.33)	*−4.25 (165)*	*.001*
Sleep onset latency, hh:mm:ss	01:31:38 (02:39:11)	00:58:37 (01:26.00)	*2.78 (165)*	*.006*
Nocturnal awakenings, n	1.89 (1.57)	1.67 (1.32)	1.81 (165)	.07
Wake after initial sleep onset, hh:mm:ss	00:48:57 (00:55:48)	00:42:01 (00:52:58)	1.27 (165)	.21
Terminal wakefulness, hh:mm:ss	00:58:11 (01:09:42)	00:39:43 (0:39:57)	*3.31 (165)*	*.001*

[a] Italicized values indicate they are statistically significant.

To measure the effect of completing Step 2 on sleep indicators, we computed the mean scores of the 7 nights before the users received personalized sleep recommendations and compared it to the mean scores of the 7 nights after they started Step 2. Mean and SD analyses among the 47 users who completed Step 2 suggest that WASO, NWAK, and TWAK decreased after Step 2, whereas TIB, TST, and sleep efficiency increased (see Table 4).

Table 4. Nocturnal sleep indicators of users completing Step 2 (n=47).

Sleep indicator	Value, mean (SD)		*t* test (*df*)	*P* value
	7 nights before Step 2	7 nights after Step 2		
Time in bed, hh:mm:ss	07:56:45 (01:18:51)	08:37:45 (00:50:21)	−5.35 (*46*)	.001
Total sleep time, hh:mm:ss	06:00:05 (01:24:29)	06:13:26 (01:30:16)	−2.02 (*46*)	.047
Sleep efficiency, %	68.47 (14.75)	72.36 (16.76)	−3.18 (*46*)	.002
Sleep onset latency, hh:mm:ss	00:59:22 (00:53:41)	01:08:50 (01:25:02)	−1.01 (*46*)	.32
Nocturnal awakenings, n	1.77 (1.24)	1.35 (0.97)	5.24 (*46*)	.001
Wake after initial sleep onset, hh:mm:ss	00:55:26 (00:52:03)	00:35:58 (00:32:57)	3.57 (*46*)	.001
Terminal wakefulness, hh:mm:ss	00:51:59 (00:42:15)	00:41:17 (00:32:22)	3.04 (*46*)	.003

Discussion

Principal Findings

Our results show, for the first time, the feasibility of using virtual agents in the context of a major health crisis to monitor insomnia symptoms and deliver assistance to the users through behavioral interventions. eHealth is a very rapidly growing field, and numerous solutions are particularly adapted to conditions such as confinement where human contacts must be limited. Several mobile apps use text-based chatbots for medical interviews, but the use of virtual agents (interacting through natural body motion and natural voice) is still sparse. We believe that these new empathic human-machine interfaces can reinforce acceptance of eHealth solutions.

More than 2000 people downloaded the KANOPEE app over the 11-day study period, with no technical errors reported by Google Play Store, indicating a higher inclusion rate than that reported in a previous study proposing digital cognitive behavioral therapy for insomnia [30]. This confirms the potential of digital technologies to provide access to clinical screening and behavioral intervention for insomnia for the general population [31]. Out of the 2069 users who downloaded KANOPEE, 1574 (76%) used it to self-evaluate their sleep. We noted that users who decided not to answer the screening interview for sleep were older and more likely to be male. This result might reflect more specific target populations such as young women—a group that is well known to report high levels of sleep complaints [32]. While the stress related to the COVID-19 confinement could explain why we obtained such a high app download rate, another potential explanation is the visual appeal of virtual agents to engage in digital interactions. Even individuals without significant sleep complaints (about 50% of the study sample) used our app and were interested in completing a sleep evaluation; this finding strongly indicates the need to develop apps for normal sleepers who want to improve their sleep hygiene and should be interpreted as a positive indicator for future sleep health campaigns.

Acceptance of the virtual agent was a major challenge in this specific context, and we obtained very good results, similar to those reported previously with outpatients in a hospital [18]. Usability and benevolence were very well ranked by the app users, which confirms the empathic dimension of our virtual agents, even on smaller devices like smartphones. This is also a positive message to reinforce the use of virtual agents in eHealth technologies, which is a growing field of interest for medicine.

In the intervention program, 21.5% of the users reporting significant sleep complaints (ie, ISI score >14) completed a daily sleep diary for more than 7 days and consented to participate in Interview 2 (ie, they completed Step 1), and 28.3% of users followed behavioral interventions and completed the sleep diary for 10 more days. Interestingly, subjects completing Step 1 significantly improved their sleep over a brief period of time (ie, 7 days). We hypothesize that filling in the sleep diary and receiving daily feedback on their sleep efficiency score helped users to adjust their sleep schedule autonomously. Another possible explanation is that their insomnia symptoms decreased naturally over time, even though a reduction of time in bed for about half an hour suggests an active change. These findings are very encouraging for the use of electronic sleep diaries to promote sleep hygiene practices, a form of low-intensity sleep health intervention that could be beneficial at the more global population level. Users who completed their personalized intervention reported an improvement in nocturnal sleep, with a reduction of nocturnal awakenings and insomnia complaints, which suggests Step 2 was beneficial for a subgroup of individuals with more significant sleep complaints. Altogether, with completion rates of 76% for the initial evaluation and 28.3% for the personalized interventions in a selected population, we believe that our results open interesting perspectives for populational interventions and mirror the proposal of Berry et al [33] to set up trials in which large-scale interventions are offered simultaneously to different subgroups of patients.

Study Limitations

Nevertheless, this study has a few limitations. First, the very peculiar period of recruitment, during the COVID-19 confinement, makes our results preliminary. Future work needs to confirm, in a more “normal period of time,” the fact that KANOPEE can help individuals improve their insomnia complaints and sleep hygiene. Second, the drop-out rate was quite high. We were unable to determine why users did not follow the program until the end, and further study is therefore needed for a precise examination of usage (eg, frequency of use and errors made) and qualitative interviews with app users to unveil the reasons why they decided to drop out of KANOPEE.

Another limitation of our study is related to the fact that we did not note an improvement among the most severe users, which shows the limitations of nonhuman interventions. Because we did not explore all the possible comorbidities, we might have proposed to some users a solution that may be unsuitable to their health problems. Future studies could use detailed interviews that could help precisely select the ideal population to receive the interventions and refer the other users directly to sleep centers.

Conclusions

Considering the above limitations, we believe that KANOPEE is a new promising tool in the field of eHhealth that could limit the number of individuals asking for consultations by general practitioners for moderate sleep complaints. Indeed, we believe this app can help in both ways: identifying individuals with insomnia complaints and providing brief and practical behavioral interventions. Further research is needed to test this app apart from the COVID-19 confinement period and on more specifically selected users.

Acknowledgments

We thank Yannick Levavasseur and Emeric Labbé for developing the app. This project was supported by the grants LABEX BRAIN (ANR-10-LABX-43), EQUIPEX PHENOVIRT (ANR-10-EQPX-12-01), and funding from the Region Nouvelle-Aquitaine (IS-OSA project, Contract No.: 18000389).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Personalized sleep recommendations and conditions.

[DOC File , 32 KB - jmir_v22i12e24268_app1.doc]

References

1. World Health Organization. Responding to community spread of COVID-19: interim guidance. World Health Organization. 2020 Mar 07. URL: https://apps.who.int/iris/handle/10665/331421
2. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
3. Morin CM, Carrier J. The acute effects of the COVID-19 pandemic on insomnia and psychological symptoms. Sleep Med 2020 Jun 06:30261-30266 [FREE Full text] [doi: 10.1016/j.sleep.2020.06.005] [Medline: 32595107]
4. Voitsidis P, Gliatas I, Bairachtari V, Papadopoulou K, Papageorgiou G, Parlapani E, et al. Insomnia during the COVID-19 pandemic in a Greek population. Psychiatry Res 2020 May 12;289:113076. [doi: 10.1016/j.psychres.2020.113076] [Medline: 33242799]
5. Stanton R, To QG, Khalesi S, Williams SL, Alley SJ, Thwaite TL, et al. Depression, anxiety and stress during COVID-19: associations with changes in physical activity, sleep, tobacco and alcohol use in Australian adults. Int J Environ Res Public Health 2020 Jun 07;17(11):4065 [FREE Full text] [doi: 10.3390/ijerph17114065] [Medline: 32517294]
6. Fagherazzi G, Goetzinger C, Rashid M, Aguayo G, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
7. Inkster B, O'Brien R, Selby E, Joshi S, Subramanian V, Kadaba M, et al. Digital health management during and beyond the COVID-19 pandemic: opportunities, barriers, and recommendations. JMIR Ment Health 2020 Jul 06;7(7):e19246 [FREE Full text] [doi: 10.2196/19246] [Medline: 32484783]
8. Sarbadhikari S, Sarbadhikari SN. The global experience of digital health interventions in COVID-19 management. Indian J Public Health 2020;64(6):117-124. [doi: 10.4103/ijph.ijph_457_20]
9. Geoffroy PA, Le Goanvic V, Sabbagh O, Richoux C, Weinstein A, Dufayet G, et al. Psychological support system for hospital workers during the Covid-19 Outbreak: rapid design and implementation of the Covid-Psy hotline. Front Psychiatry 2020 May 28;11:511. [doi: 10.3389/fpsyt.2020.00511] [Medline: 32670100]
10. Taylor CB, Fitzsimmons-Craft EE, Graham AK. Digital technology can revolutionize mental health services delivery: The COVID-19 crisis as a catalyst for change. Int J Eat Disord 2020 Jul;53(7):1155-1157 [FREE Full text] [doi: 10.1002/eat.23300] [Medline: 32449523]
11. Your Safe Space In This Difficult Time. WYSA. 2020. URL: https://www.wysa.io/ [accessed 2020-12-07]
12. Ritterband LM, Thorndike FP, Gonder-Frederick LA, Magee JC, Bailey ET, Saylor DK, et al. Efficacy of an internet-based behavioral intervention for adults with insomnia. Arch Gen Psychiatry 2009 Jul;66(7):692-698 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.66] [Medline: 19581560]

13. Espie C, Kyle S, Williams C, Ong J, Douglas N, Hames P, et al. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. Sleep 2012 Jun 01;35(6):769-781 [FREE Full text] [doi: 10.5665/sleep.1872] [Medline: 22654196]
14. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - A systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
15. Philip P, Bioulac S, Sauteraud A, Chaufton C, Olive J. Could a virtual human be used to explore excessive daytime sleepiness in patients? Presence: Teleoperators and Virtual Environments 2014 Nov 01;23(4):369-376. [doi: 10.1162/pres_a_00197]
16. Philip P, Micoulaud-Franchi JA, Sagaspe P, de Sevin E, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Feb 16;7:42656 [FREE Full text] [doi: 10.1038/srep42656] [Medline: 28205601]
17. Auriacombe M, Moriceau S, Serre F, Denis C, Micoulaud-Franchi J, de Sevin E, et al. Development and validation of a virtual agent to screen tobacco and alcohol use disorders. Drug Alcohol Depend 2018 Dec 01;193:1-6. [doi: 10.1016/j.drugalcdep.2018.08.025] [Medline: 30321739]
18. Philip P, Dupuy L, Auriacombe M, Serre F, de Sevin E, Sauteraud A, et al. Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. NPJ Digit Med 2020 Jan 07;3:2 [FREE Full text] [doi: 10.1038/s41746-019-0213-y] [Medline: 31934646]
19. Dupuy L, Micoulaud-Franchi JA, Philip P. Acceptance of virtual agents in a homecare context: Evaluation of excessive daytime sleepiness in apneic patients during interventions by continuous positive airway pressure (CPAP) providers. J Sleep Res 2020 Jun 18:e13094. [doi: 10.1111/jsr.13094] [Medline: 32557996]
20. Santé Publique France - Observatoire Géodes. URL: https://geodes.santepubliquefrance.fr/#bbox=-1366851,6775601,3198114,1847697&c=indicator&f=0&i=covid_hospit.hosp&s=2020-03-18&t=a01&view=map2 [accessed 2020-12-09]
21. Bastien C, Vallières A, Morin C. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Medicine 2001;2(4):297-307 [FREE Full text] [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
22. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
23. Kanopee project - SANPSY Lab. YouTube. 2020 Jun 25. URL: https://www.youtube.com/watch?v=y3QfgNvQfr0&feature=youtu.be [accessed 2020-12-07]
24. Morin C. Insomnia: Psychological assessment and management. In: APA PsycInfo. New York, US: Guilford Press; 1993.
25. Buysse D, Ancoli-Israel S, Edinger J, Lichstein K, Morin CM. Recommendations for a standard research assessment of insomnia. Sleep 2006 Sep;29(9):1155-1173. [doi: 10.1093/sleep/29.9.1155] [Medline: 17040003]
26. Ewing JA. Detecting Alcoholism: the CAGE Questionnaire. JAMA 1984 Oct 12;252(14):1905. [doi: 10.1001/jama.1984.03350140051025]
27. Etter J, Le Houezec J, Perneger TV. A self-administered questionnaire to measure dependence on cigarettes: the cigarette dependence scale. Neuropsychopharmacology 2003 Feb;28(2):359-370. [doi: 10.1038/sj.npp.1300030] [Medline: 12589389]
28. Micoulaud-Franchi JA, Sauteraud A, Olive J, Sagaspe P, Bioulac S, Philip P. Validation of the French version of the Acceptability E-scale (AES) for mental E-health systems. Psychiatry Res 2016 Mar 30;237:196-200. [doi: 10.1016/j.psychres.2016.01.043] [Medline: 26809367]
29. Tariman JD, Berry DL, Halpenny B, Wolpin S, Schepp K. Validation and testing of the Acceptability E-scale for web-based patient-reported outcomes in cancer care. Appl Nurs Res 2011 Feb;24(1):53-58 [FREE Full text] [doi: 10.1016/j.apnr.2009.04.003] [Medline: 20974066]
30. Espie CA, Emsley R, Kyle SD, Gordon C, Drake CL, Siriwardena AN, et al. Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. JAMA Psychiatry 2019 Jan 01;76(1):21-30 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.2745] [Medline: 30264137]
31. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
32. Zhang B, Wing YK. Sex differences in insomnia: a meta-analysis. Sleep 2006 Jan;29(1):85-93. [doi: 10.1093/sleep/29.1.85] [Medline: 16453985]
33. Berry SM, Connor JT, Lewis RJ. The platform trial: an efficient strategy for evaluating multiple treatments. JAMA 2015 Apr 28;313(16):1619-1620. [doi: 10.1001/jama.2015.2316] [Medline: 25799162]

Abbreviations

AES: Acceptability of e-health scale
CAGE: Cut-down, annoyed, guilty, eye-opener questionnaire
CDS-5: Cigarette Dependence Scale (5-item) questionnaire
CNIL: Commission nationale de l'informatique et des libertés

ECAs: embodied conversational agents
ETQ: ECA trust questionnaire
GDPR: General Data Protection Regulation
ISI: insomnia severity index
NWAK: number of awakenings
SOL: sleep onset latency
TIB: time spent in bed
TST: total sleep time
TWAK: terminal wakefulness
WASO: wake after initial sleep onset

Edited by G Eysenbach; submitted 11.09.20; peer-reviewed by J Manning, N Khalili-Mahani; comments to author 13.10.20; revised version received 26.10.20; accepted 30.11.20; published 18.12.20.

Please cite as:
Philip P, Dupuy L, Morin CM, de Sevin E, Bioulac S, Taillard J, Serre F, Auriacombe M, Micoulaud-Franchi JA
Smartphone-Based Virtual Agents to Help Individuals With Sleep Concerns During COVID-19 Confinement: Feasibility Study
J Med Internet Res 2020;22(12):e24268
URL: http://www.jmir.org/2020/12/e24268/
doi:10.2196/24268
PMID:33264099

Original Paper

Reduction of COVID-19 Incidence and Nonpharmacologic Interventions: Analysis Using a US County–Level Policy Data Set

Senan Ebrahim[1,2*], PhD; Henry Ashworth[1,2*], MPH; Cray Noah[1,2], BSc; Adesh Kadambi[2,3], BEng; Asmae Toumi[4], BSc; Jagpreet Chhatwal[1,4], PhD

[1]Harvard Medical School, Boston, MA, United States
[2]Hikma Health, San Jose, CA, United States
[3]University of Toronto, Toronto, ON, Canada
[4]Massachusetts General Hospital, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Senan Ebrahim, PhD
Harvard Medical School
25 Shattuck St
Boston, MA, 02115
United States
Phone: 1 617 432 1000
Email: senan@hikmahealth.org

Abstract

Background: Worldwide, nonpharmacologic interventions (NPIs) have been the main tool used to mitigate the COVID-19 pandemic. This includes social distancing measures (closing businesses, closing schools, and quarantining symptomatic persons) and contact tracing (tracking and following exposed individuals). While preliminary research across the globe has shown these policies to be effective, there is currently a lack of information on the effectiveness of NPIs in the United States.

Objective: The purpose of this study was to create a granular NPI data set at the county level and then analyze the relationship between NPI policies and changes in reported COVID-19 cases.

Methods: Using a standardized crowdsourcing methodology, we collected time-series data on 7 key NPIs for 1320 US counties.

Results: This open-source data set is the largest and most comprehensive collection of county NPI policy data and meets the need for higher-resolution COVID-19 policy data. Our analysis revealed a wide variation in county-level policies both within and among states ($P<.001$). We identified a correlation between workplace closures and lower growth rates of COVID-19 cases ($P=.004$). We found weak correlations between shelter-in-place enforcement and measures of Democratic local voter proportion (R=0.21) and elected leadership (R=0.22).

Conclusions: This study is the first large-scale NPI analysis at the county level demonstrating a correlation between NPIs and decreased rates of COVID-19. Future work using this data set will explore the relationship between county-level policies and COVID-19 transmission to optimize real-time policy formulation.

(J Med Internet Res 2020;22(12):e24614) doi:10.2196/24614

KEYWORDS

communicable diseases; COVID-19; data set; pandemic; policy; public health; data; intervention; effectiveness; incidence; time series

Introduction

In the absence of effective vaccines or therapeutics targeting SARS-CoV-2, nonpharmacologic interventions (NPIs) have been the only effective measures for containing the current COVID-19 pandemic [1-4]. NPIs are grouped into social distancing measures and contact tracing measures [5]. Examples of social distancing measures include closing businesses, closing schools, and quarantining symptomatic persons [6]. Contact tracing involves tracking and following exposed individuals, which requires both testing capabilities and infrastructure to execute [7]. Systematic reviews and modeling studies have demonstrated that each of these NPIs have a variable impact on respiratory virus transmission, depending on how and when they are deployed [1,5,7].

While China was initially slow to implement NPIs, models have shown that China’s social distancing measures were sufficient to control COVID-19 [1,2]. Local governments across China were integral in the successful implementation of NPIs, including diagnostic testing and enforcing social distancing [2]. NPIs have similarly been found to effectively limit COVID-19 across 11 countries in Europe [4], especially Italy [3]. However, the United States has implemented NPIs more variably, which may be related to the higher rates of transmission [8].

In the United States, the first known cases of COVID-19 were reported on January 20, 2020 [9]. However, it was not until March 2020 that individual states responded with NPI policies, and as of April 20, 2020, multiple states still had not implemented stay-at-home orders [10,11]. The limited coordination of national, state, and local responses to COVID-19 led to substantial variation in NPIs at the county level across the United States. As the pandemic continued into late May and early June, states began to roll back measures with incongruent reopening plans. These reopening plans again differed by location and by NPI policy type [9]. As a result, the United States has uniquely experienced wide variations in NPI policy, both geographically and temporally. County-level political party alignments may be relevant to the formulation of local COVID-19 NPI policies; this question has not yet been rigorously assessed. Partisanship in county-level policy has significant ramifications for how policymakers engage stakeholders to realistically implement local policy.

As of November 25, 2020, there are 12,838,102 confirmed COVID-19 cases and 262,847 deaths in the United States [8], both higher totals than those of any other country in the world [8,12]. Preliminary data on social distancing at the state level [13] and limited data on shelter-in-place orders at the county level from bordering communities in Illinois and Iowa [14] have shown that NPIs can be effective, particularly when implemented in a timely manner. However, there remains a need to better understand the effects of NPIs on COVID-19 transmission dynamics at a national scale on the granular county level.

The first objective of this study was to describe the motivation and novel methodology for creating the first large-scale county-level NPI policy data set in the United States. The second objective was to highlight initial findings from this data set to demonstrate its utility for much-needed local NPI analyses. A unique standardized crowdsourcing methodology was deployed to collect time-series data on 7 key NPIs for 1320 US counties. This novel data set was then mined for correlations in combination with publicly available COVID-19 case data, reproduction number (R_t) estimates, and political demographics at the county level. This exploratory analysis illustrates the utility of county-level NPI implementation and analysis in the United States, particularly with the novel data set described herein.

Methods

Data Collection

A novel crowdsourcing methodology was implemented by Hikma Health to collect COVID-19 policy data for 1320 US counties from March to July 2020. The data set covers 7 distinct NPI policies, including the most widely deployed and accepted NPIs. For each county, trained volunteers reported a binarized policy status for each NPI policy indicating if/when the following NPI policies were first implemented in each county, along with a timestamp and a reference URL: (1) closure of nonessential workplaces, (2) shelter-in-place/stay-at-home orders, (3) enforcement of shelter-in-place/stay-at-home orders, (4) size restrictions on public gatherings, (5) school closures, (6) public transit closures, and (7) publicly available testing. Table 1 provides a full description of the variables included in the data set. The data set also includes a second timestamp and reference URL for if/when each of the same 1320 US counties terminated the following 2 NPI policies: nonessential workplace closures and shelter-in-place/stay-at-home orders. Given the limits in data collection capacity, these 2 NPI policies were prioritized over the other 5 NPI policies for observation at a second timepoint, as we hypothesized they had a relatively higher likelihood of changing in May-July 2020.

Table 1. Variables in the Hikma Health data set of county-level nonpharmacologic intervention policies.

Variable name	Description
fips	County FIPS[a] geographic code (unique identifier)
testing	Binary coding whether COVID-19 testing is publicly available in the county to any resident without physician referral needed
school	Binary coding whether all schools are closed in the county
shelter	Binary coding whether the county has an active shelter-in-place order, publicly announced by any county official
shelter_enforcement	Binary coding whether the shelter-in-place order is being enforced in the county with fines or other penalties
work	Binary coding whether all "nonessential workplaces" are closed in the county, with any local definition of "nonessential"
event	Binary coding whether public events and gatherings larger than a particular size N are restricted, for any N>1
transport	Binary coding whether any public transportation has been closed down for any public bus, train, shuttle, or ferry routes
X_date	For each policy binary X, the date on which it was first implemented
X_URL	For each policy binary X, the source URL with evidence of the nature and date of the policy
updated	The timestamp for when this data was entered

[a] FIPS: Federal Information Processing Standard.

Because each US county reports its standing COVID-19 NPI policies differently—from county websites to local news outlets and official social media channels—the data collection process cannot be automated and instead requires human review and discernment. From March to July of 2020, 104 volunteers, consisting mostly of health-related graduate degree students and medical professionals, were recruited through COVID-19 project postings, outreach groups, and institutional listservs. Each volunteer was remotely trained to use the same 7-step standard operating procedure to research and collect the aforementioned NPI data on 1320 US counties through standardized online forms, effectively transforming the convoluted county policy landscape into an organized NPI data set with binary yes/no and interval date variables. The free and open-source data set contains corresponding URL references on all counties for quality assurance [15].

In assigning counties to volunteers, we initially prioritized population and then transitioned to COVID-19 incidence as the pandemic unfolded. Specifically, we sourced data for the 500 most populous US counties and then used dynamic 4-day incidence rate calculations to prioritize the remaining counties in real time. For the first 100 counties, every policy and its implementation date was validated manually by double-checking the website URL from which the information was sourced. The same extraction process was repeated for the subsequent 1220 counties, with validation of URLs for an additional randomly selected 10% of completed counties, rendering a volunteer coding accuracy rate over 99%.

Statistical Analysis

In this study, we conducted time-series correlational analyses combining our county-level NPI data for 1320 US counties with multiple data sources, including daily county COVID-19 cases and deaths sourced from The New York Times. We also assessed correlations with COVID-19 effective R_t estimates from the RT Live project, and political demographics at the county level from the Kaiser Family Foundation and the MIT Election Data and Science Lab.

For optimal visualization and temporally focused analysis on current policies, we constructed a consolidated version of the data set as follows: for each NPI policy in each county, all observations within the last 24 hours were pooled; the mean was calculated for each binary, and binaries above 0.5 were considered positive, while binaries less than or equal to 0.5 were considered negative; the latest date and URL reported were chosen to represent that policy; lastly, a Policy Strength Index (PSI) was calculated as a linear sum of all 7 NPI policy binary variables, with 7 being the maximum and 0 being the minimum possible value.

We then mined this consolidated data set for correlations and distribution differences using standard statistical tests including *T* tests and chi-square tests with a Bonferroni correction applied for multiple hypothesis testing. All analyses were conducted in Python notebooks that are available open source for review and global use under the Apache 2.0 license.

Results

Data Set Construction and Access

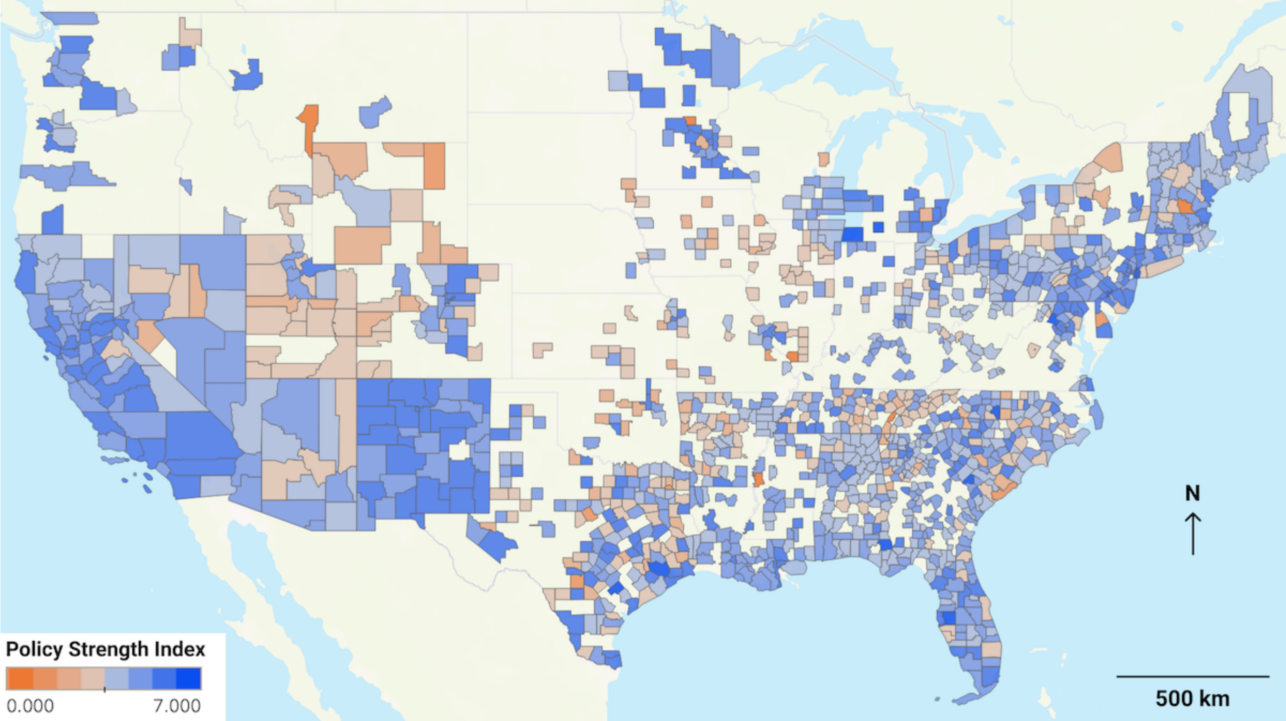
The full county NPI data set, hereon referred to as the "all policies" data set, yielded 2704 observations of NPI policies described in Table 1 in 1320 counties from all 50 states in the United States. We analyzed the all policies data set (n=2704) as well as the consolidated current version of the data set, hereon referred to as the "current policies" data set (n=1320), containing only the most recent timestamp for county NPIs as described in the *Methods* section. These versions are referred to as the "all_county_ policies" and "county_policies" files, respectively, on GitHub, where they can be freely accessed by the global public, along with reference documentation [15]. The data set is available as both CSV (comma-separated values) and JSON (JavaScript Object Notation) files, indexed by US county FIPS (Federal Information Processing Standard) geographic codes. Each of these files include the binary NPI variables with accompanying interval date variables, corresponding reference URLs, and the timestamps of when the data were collected. The

CSV files used for all analyses presented in this paper are included in Multimedia Appendix 1.

Policy and Case Correlations

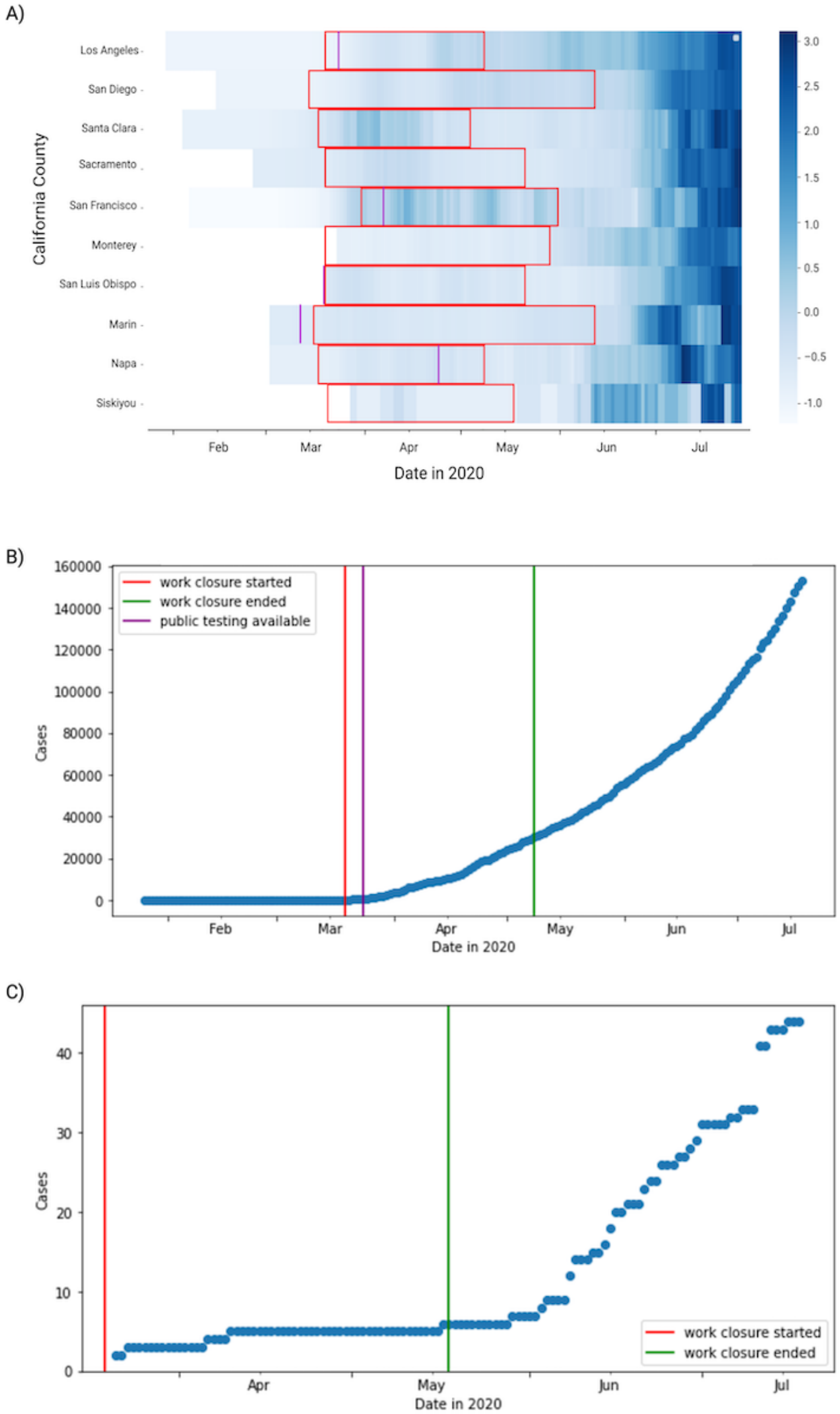
In the all policies data set, there was a strong positive correlation between nonessential workplace closure and shelter-in-place orders at the county level with a Pearson R of 0.835. All other correlations between individual policies were weak in the all policies data set, with an absolute value of R<0.3. In the current policies data set, the correlation between nonessential workplace closure and shelter-in-place orders weakened to a Pearson R of 0.144, generally in association with the reopening of workplaces without lifting shelter-in-place orders in May-June 2020. In the current policies data set, all other correlations were less than 0.1. Of all policies tested for interstate differences by a one-way ANOVA (analysis of variance), only school closure was not significantly different at a Bonferroni corrected α level of .0071 (*P*=.06); for all other policies *P*<.001. States also exhibited varying degrees of overall intrastate/intercounty variability in PSI, as illustrated in Figure 1.

Figure 1. Map showing that county policies are uncorrelated and highly variable between and within states. Policy Strength Index (PSI) for the current policies data set mapped by county indicates that across states, there was wide variability in policy implementation on July 15, 2020. The PSI is calculated as the linear sum of the 7 binary nonpharmacologic intervention policy variables in each county as described in the Methods section.



We also observed that within states the variability in county NPIs informs the visualization of case growth rates, as shown in Figure 2. For example, in California, the variability in county NPI policies informed the interpretation of timelines showing the 7-day simple moving averages of new cases in each county (Figure 2A). As a general trend, case growth was minimal during the period of workplace closure, highlighted for each county in Figure 2A in the red window, while case growth increased significantly 2-8 weeks after the end of workplace closure. For example, in Los Angeles County, California (shown in Figure 2B), case growth accelerated following the end of the work closure policy; additionally, the start of public testing coincided with the inflection point of case growth. In Siskiyou County, California (shown in Figure 2C), case growth similarly accelerated following the end of workplace closure.

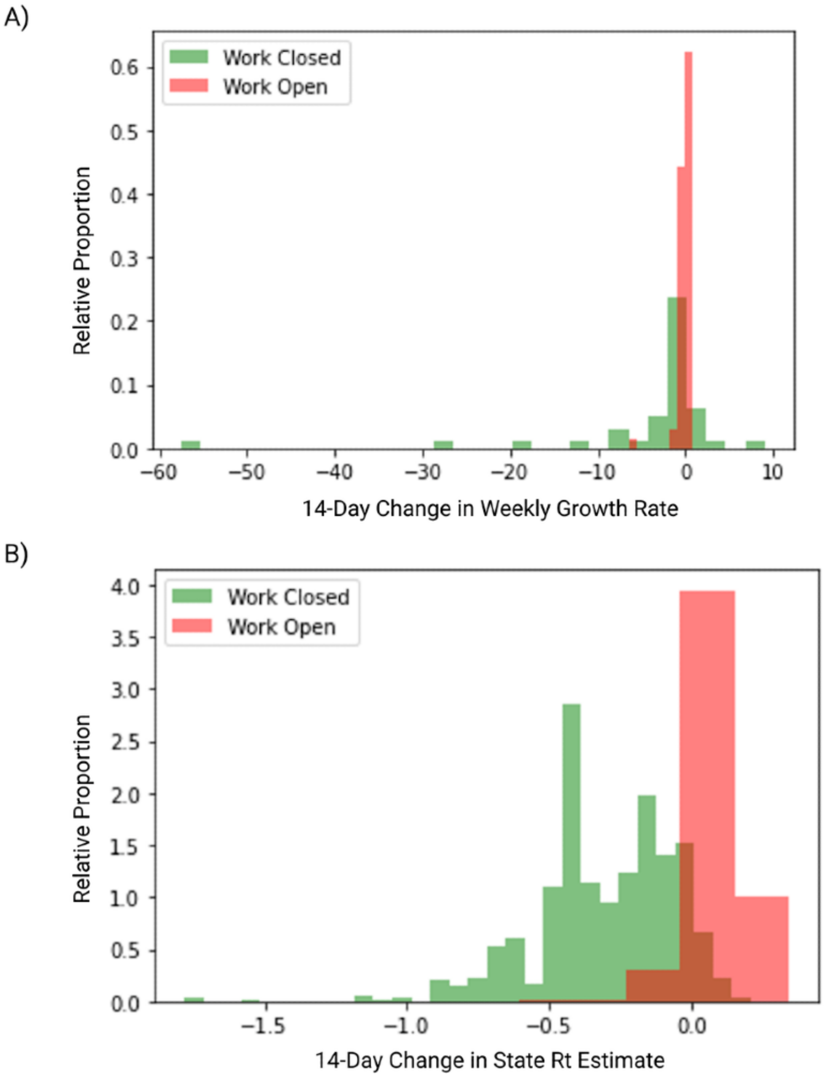
Figure 2. Variability in California county policies in time and in relation to caseloads. (A) Time-series heatmap of the 7-day simple moving average of daily cases plotted for 10 representative California counties in order of descending population, all normalized by county. The red rectangle demarcates the period for which the county had nonessential workplaces closed. The purple line indicates if and when public testing became available. (B) Plot of the total number of county cases as a timeline with county-level nonessential work closure policy and public testing times labeled for Los Angeles County. (C) Plot of the total number of county cases as a timeline with county-level nonessential work closure policy times labeled for Siskiyou County.



Policies and Case Growth

As summarized in Figure 3, we assessed whether there are significant differences in the change in weekly case growth in counties where workplaces closed versus those that remained open in the all policies data set. Figure 3A shows a histogram illustrating change in weekly case growth rate from the week preceding the date of workplace closure to 14 days later for counties with open versus closed nonessential workplaces. Weekly case growth rates decreased over the 14 days following a workplace NPI by larger magnitudes for counties that had such nonessential workplace closure policies as compared to those that had openings (*P*=.004; Figure 3A). We also assessed whether there was a significant difference in statewide R_t. This effect of decreasing case propagation following workplace closure was similarly observed in state-level estimates of R_t 14 days after the workplace policy (from RT Live) (*P*<.001; Figure 3B). The effect for R_t was recapitulated at 28 days after the workplace policy (*P*<.001). These effects are statistically significant at our Bonferroni corrected α level of .0071.

Figure 3. Counties with closed nonessential workplaces and significant declines in growth rate and reproduction number (R_t) following workplace closure. (A) A histogram of the change in weekly case growth rate from the week preceding the date of workplace closure to 14 days later for counties with open versus closed nonessential workplaces. (B) A histogram of the change in RT Live estimates of R_t by state from the date of the workplace policy to 14 days later for counties with open versus closed nonessential workplaces.

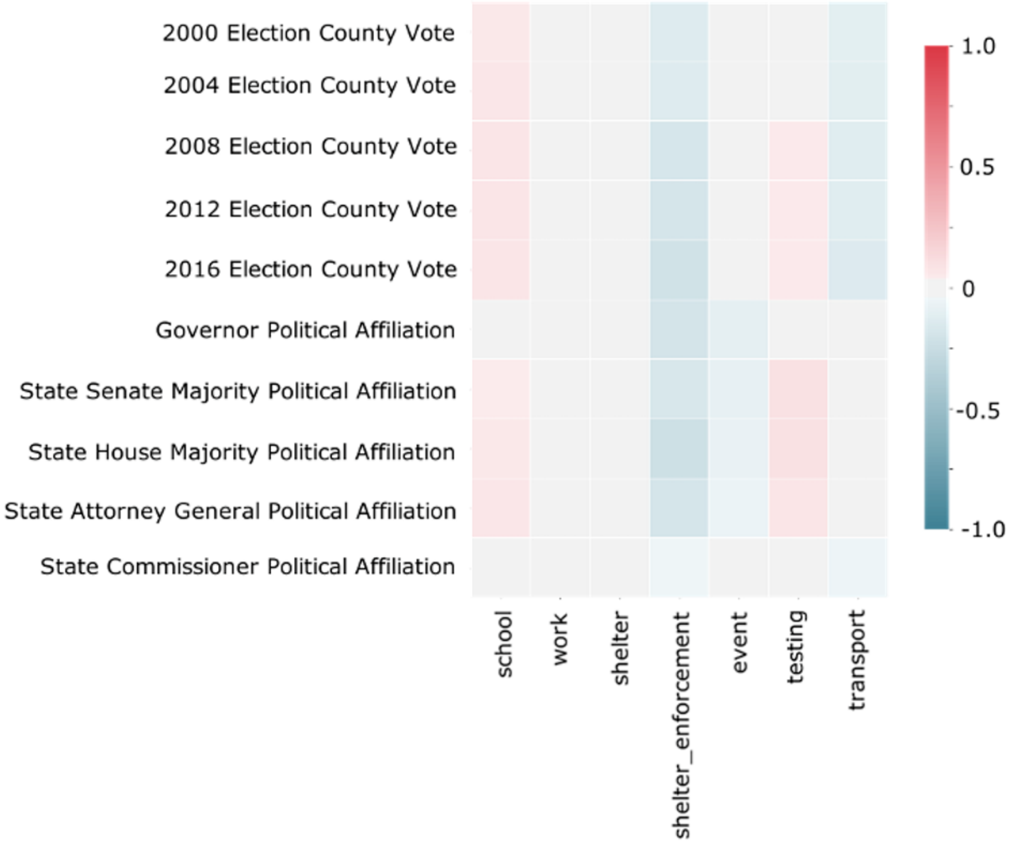


The start date of public testing varied in each county both on the absolute date timeline and the timeline relative to other NPIs. To assess whether counties that had free public testing before a workplace policy observation were more likely to have had a workplace closure for that observation, counties were grouped based on their timelines into the categories of no testing, testing after workplace closure, and testing before workplace closure. A chi-square test for significance found no significant difference in workplace openings versus workplace closures for these categories (*P*=.08).

Policies and Political Alignment Correlations

We observed weak correlations between the political parties of local leadership and electorate and the policy of shelter-in-place enforcement. In the all policies data set, shelter-in-place enforcement was weakly correlated with Democratic party State House leadership (R=0.22) and Democratic voting proportion in the 2016 presidential election (R=0.21); all other correlations had an absolute value of R<0.2 (Figure 4). No correlations greater than 0.2 were observed between county political parties and the dates at which a positive intervention was initially made for a particular NPI policy.

Figure 4. Weak correlation between county policies and county political parties. Correlogram of county-level nonpharmacologic interventions (NPIs) and political election patterns for state and national government for the all policies data set. A positive correlation shown in red indicates a positive correlation between a Republican-oriented election pattern and an NPI.



Discussion

Principal Results

The most effective measures for countries around the world preventing COVID-19 have been NPIs [1-4]. However, due to policy variations across levels of government, the United States has not demonstrated the same success with NPIs [8]. While some research has been published looking at small collections of neighboring counties, there has not been a comprehensive analysis of NPIs across a wider range of counties or any local level more granular than the state level in the United States [14]. In this study, we identify associations between county level NPI policies and COVID-19 transmission from March 1 to July 20, 2020, using a novel crowdsourced county NPI policy data set.

The first objective of this project was to construct a comprehensive data set that has a broad representation of different NPI policies; has a diverse representation of counties both in geographic distribution and population; and incorporates changes in NPIs over time. Overall, this study rendered a highly granular NPI policy data set with time-series data for 1320 counties from all 50 states in the United States. While there have been comprehensive data sets reviewing NPI policies at the state level [10,11] and some data generated for small clusters of neighboring counties (<100 counties) [14], this novel data set is the most comprehensive publicly available county data set to date. We have openly shared this data set with the aim that it be used by researchers around the United States and the world to further analyze the correlates of NPIs and various COVID-19 epidemiological outcomes. Future research could include a finer-grained analysis and modeling of the relationships between COVID-19 cases, deaths, and NPI policies, as well as the explorations of the relationship between NPIs and outcomes including economic status, health care utilization, and social inequities.

The second objective of this project was to analyze this data set and report preliminary findings. We found that across the United States, there was significant variability in NPI policy implementation among counties, both within and among states. NPI policies themselves for any given county are largely uncorrelated, with the singular exception of workplace closure and shelter-in-place NPIs. This finding supports previous research that has also found wide variation in policies across states, which was associated with statistically significant differences in rates of COVID-19 transmission [13]. While no causal link has been demonstrated, these strong correlations are further supported by international studies that have compared similarly sized territories within countries [2,3].

We also found that there was significant variation in NPI policies across counties within states. We displayed 10 representative California counties as an example, which show the relationship between the rate of COVID-19 cases and timing of when individual California counties closed workplaces, reopened workplaces, and started public testing. These results show that the implementation of NPI policies varied when they were enacted in response to COVID-19 case rates across counties. As a consistent trend, we observed that for both heavily populated urban counties like Los Angeles County (10 million residents) and sparsely populated rural counties like Siskiyou County (44,000 residents), the end of a workplace closure appears to precede an increase in cases by 2-6 weeks. This heatmap timeline could be confounded by when public testing started in each county and the rate at which counties were able to test. While we have recorded the date at which public testing became available in each county, as displayed in Figure 2, the rate of testing by county remains generally unknown.

The final and most notable finding of this analysis is that there are statistically significant positive correlations between county workplace closures and decreases in subsequent COVID-19 case growth, as shown in Figure 3. We found that across all counties, those with work closures had significantly lower rates of subsequent COVID-19 case growth compared to counties that did not. We emphasize that these results do not in any way demonstrate causation but rather a temporally informative correlation between NPI policy and COVID-19 rates. Nevertheless, these findings align with others highlighting the relationship between increased NPI policies and subsequently declining growth rates of COVID-19 at the county level [14]. Our Hikma Health data set also has the potential to be used in further temporal modeling, particularly to predict R_t for counties across the United States. Our data set particularly lends itself to a clustering analysis to assess the relationship between NPI policies and cases in demographically similar counties, which will be helpful to local policymakers.

Current research on political affiliation and attitudes toward COVID-19 policies has shown that differences between Democrats and Republicans are more significant than differences across race or gender [16]. Reports from the Pew Research Center have shown that Democrats are more likely to see COVID-19 as a serious threat [17]. In order to understand the political landscape in which these policies are being implemented, we analyzed both the political parties in charge of counties and the political electorate in each county. As shown in Figure 4, we found a relatively weak correlation between Democratic county governments and electorates and shelter-in-place policy enforcement (no significant correlations). However, our finding of relatively limited differences by political party is consistent with a generally high public support across the political spectrum for NPI policies to reduce COVID-19 spread [18]. These preliminary results should be further analyzed to understand the relationship between political party and NPI policy choices.

Limitations

There are a number of important limitations to note about our study. Firstly, our data collection, while rigorous, is affected by a number of factors both inherent to the study and external that could skew outcomes. Even though data collectors were well trained and used standardized methods, the estimated date of policy changes could be highly variable, particularly for counties with limited or conflicting information available online. Thus, an important next step will be to fully validate this Hikma Health data set by double coding all 1320 counties and subsequent reconciliation of discordant datapoints. In addition, external and contemporaneous factors such as rates of testing and the degree to which the public actually adhere to NPI guidelines are not addressed by our binarized data set. In subsequent analyses to address these factors, we aim to integrate testing rates and mobility data as these data sets become available.

Secondly, the Hikma Health data set comprises 2 timepoints for 2 of the 7 NPI policies originally assessed. The analyses presented in Figures 2 and 3 are subject to the caveat that changes in the other 5 unrecorded NPI policies may theoretically confound any associations. Of these NPI policies, enforcement of shelter-in-place/stay-at-home orders, school closures, and publicly available testing were unlikely to change before August 2020, whereas size restrictions on public gatherings and public transit closures may have been reversed and therefore might be more significant confounds.

Finally, our analysis identifies correlations in the data set without any implication of causality. In order to establish causation, NPI policies would have to be implemented as a coordinated randomized controlled trial across counties, which is unlikely. In lieu of such a study, our group and others will build temporal predictive models using this data set to test the potential effects of NPI policies.

Despite these limitations, our study is the first to identify correlations between county-level NPI policies and subsequent COVID-19 growth rates across the United States, including over 1000 counties from all 50 states. Our novel data set enabled us to consistently describe correlations for counties across the United States, compared to previous studies conducted on a more limited and thus less representative scale [13,14]. Our data collection methodology allowed for the evaluation and validation of data across geography and time. As the pandemic continues, future research should continue to investigate the relationship between NPI policies, COVID-19 case rates, and the factors that may influence implementation such as political affiliation, culture, and social structures.

Conclusion

As COVID-19 cases continue to climb across the United States, we anticipate that local leadership at the state and county levels will need to devise informed and relevant policies to limit local spread. Our findings suggest that there is substantial variation in NPI implementation and termination at the county level, both between and within states, reflecting an inconsistent policy approach. We also found positive correlations between implementing a workplace closure NPI and lower future rates of COVID-19, supporting previous national and international studies suggesting that NPI policies like workplace closure reduce COVID-19 transmission [1-4,13,14]. Taken together, this growing body of literature suggests that NPI policies at multiple levels, and especially at the local level of the county, play a role in limiting the effects of the COVID-19 pandemic.

Acknowledgments

This work was supported by the National Science Foundation (2035361 to JC), the National Institutes of Health (F31NS105161, T32GM007753, and T32MH020017 to SE), the Harvard Medical Scientist Training Program (to SE), and the Paul & Daisy Soros Fellowship for New Americans (to SE). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation or the National Institutes of Health.

Authors' Contributions

SE and HA were responsible for study concept and design, with help from JC. Data acquisition was conducted by CN, SE, AK, and HA. SE, AK, and AT analyzed and interpreted the data. HA and SE drafted the manuscript. JC, HA, SE, and AT critically revised the paper for important intellectual content.

Conflicts of Interest

JC reports receiving consulting fees from Gilead Sciences, Merck, and Value Analytics Labs, all unrelated to this work. SE, HE, CN, AK, and AT have no relevant conflicts of interest. The authors will not receive any payment or reward of any kind for writing this paper.

Multimedia Appendix 1

HikmaHealthCountyData.zip contains two CSV data files analyzed in this paper: 'all_county_policies.csv' with all policies reported for each county, and 'county_policies.csv' with a synthesized policy summary for each county (generated by process described in the manuscript).

[ZIP File (Zip Archive), 372 KB - jmir_v22i12e24614_app1.zip]

References

1. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science 2020 Jun 26;368(6498):1481-1486 [FREE Full text] [doi: 10.1126/science.abb8001] [Medline: 32350060]
2. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature 2020 Sep 04;585(7825):410-413. [doi: 10.1038/s41586-020-2293-x] [Medline: 32365354]
3. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020 Jun;26(6):855-860 [FREE Full text] [doi: 10.1038/s41591-020-0883-7] [Medline: 32322102]
4. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Imperial College COVID-19 Response Team, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020 Aug 08;584(7820):257-261. [doi: 10.1038/s41586-020-2405-7] [Medline: 32512579]
5. Fong MW, Gao H, Wong JY, Xiao J, Shiu EY, Ryu S, et al. Nonpharmaceutical Measures for Pandemic Influenza in Nonhealthcare Settings-Social Distancing Measures. Emerg Infect Dis 2020 May;26(5):976-984 [FREE Full text] [doi: 10.3201/eid2605.190995] [Medline: 32027585]
6. Markel H, Lipman HB, Navarro JA, Sloan A, Michalsen JR, Stern AM, et al. Nonpharmaceutical interventions implemented by US cities during the 1918-1919 influenza pandemic. JAMA 2007 Aug 08;298(6):644-654. [doi: 10.1001/jama.298.6.644] [Medline: 17684187]
7. Klinkenberg D, Fraser C, Heesterbeek H. The effectiveness of contact tracing in emerging epidemics. PLoS One 2006 Dec 20;1:e12 [FREE Full text] [doi: 10.1371/journal.pone.0000012] [Medline: 17183638]
8. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/S1473-3099(20)30120-1]
9. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Washington State 2019-nCoV Case Investigation Team. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med 2020 Mar 05;382(10):929-936 [FREE Full text] [doi: 10.1056/NEJMoa2001191] [Medline: 32004427]
10. State Action on Coronavirus (COVID-19). NCSL. 2020 Dec 16. URL: https://www.ncsl.org/research/health/state-action-on-coronavirus-covid-19.aspx [accessed 2020-11-25]
11. Mervosh S, Lu D, Swales V. See Which States and Cities Have Told Residents to Stay at Home. The New York Times. 2020 Apr 20. URL: https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html [accessed 2020-11-25]
12. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-11-25]
13. Siedner MJ, Harling G, Reynolds Z, Gilbert RF, Haneuse S, Venkataramani AS, et al. Correction: Social distancing to slow the US COVID-19 epidemic: Longitudinal pretest-posttest comparison group study. PLoS Med 2020 Oct 6;17(10):e1003376 [FREE Full text] [doi: 10.1371/journal.pmed.1003376] [Medline: 33022016]

14. Lyu W, Wehby GL. Comparison of Estimated Rates of Coronavirus Disease 2019 (COVID-19) in Border Counties in Iowa Without a Stay-at-Home Order and Border Counties in Illinois With a Stay-at-Home Order. JAMA Netw Open 2020 May 01;3(5):e2011102 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.11102] [Medline: 32413112]
15. Noah C, Ebrahim S, Ashworth H, Ebrahim A, Kadambi A, Pattilachan T. Covid-19 County Level Policy Data. Hikma Health / GitHub. 2020. URL: https://github.com/hikmahealth/covid19countymap [accessed 2020-11-25]
16. Van Green T, Tyson A. 5 facts about partisan reactions to COVID-19 in the U.S. Pew Research Center. 2020 Apr 2. URL: https://www.pewresearch.org/fact-tank/2020/04/02/5-facts-about-partisan-reactions-to-covid-19-in-the-u-s/ [accessed 2020-11-25]
17. Doherty C, Kiley J, Daniller A, Jones B, Hartig H, Dunn A. Republicans, Democrats move even further apart in coronavirus concerns. Pew Research Center. 2020 Jun 25. URL: https://www.pewresearch.org/politics/2020/06/25/republicans-democrats-move-even-further-apart-in-coronavirus-concerns/ [accessed 2020-11-25]
18. Czeisler ME, Tynan MA, Howard ME, Honeycutt S, Fulmer EB, Kidder DP, et al. Public Attitudes, Behaviors, and Beliefs Related to COVID-19, Stay-at-Home Orders, Nonessential Business Closures, and Public Health Guidance - United States, New York City, and Los Angeles, May 5-12, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):751-758 [FREE Full text] [doi: 10.15585/mmwr.mm6924e1] [Medline: 32555138]

Abbreviations

ANOVA: analysis of variance
CSV: comma-separated values
FIPS: Federal Information Processing Standard
JSON: JavaScript Object Notation
NPI: nonpharmaceutical intervention
PSI: Policy Strength Index
R_t: reproduction number

Edited by G Eysenbach; submitted 27.09.20; peer-reviewed by P Postigo-Martin, G Long; comments to author 25.11.20; accepted 08.12.20; published 21.12.20.

Please cite as:
Ebrahim S, Ashworth H, Noah C, Kadambi A, Toumi A, Chhatwal J
Reduction of COVID-19 Incidence and Nonpharmacologic Interventions: Analysis Using a US County–Level Policy Data Set
J Med Internet Res 2020;22(12):e24614
URL: http://www.jmir.org/2020/12/e24614/
doi:10.2196/24614
PMID:33302253

Original Paper

Social Capital–Accrual, Escape-From-Self, and Time-Displacement Effects of Internet Use During the COVID-19 Stay-at-Home Period: Prospective, Quantitative Survey Study

Cecilia Cheng[1], PhD; Yan-Ching Lau[2], BSc; Jeremy W Luk[3], PhD

[1]Social and Health Psychology Laboratory, Department of Psychology, The University of Hong Kong, Hong Kong, Hong Kong
[2]Division of Psychology and Language Sciences, University College London, London, United Kingdom
[3]National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD, United States

Corresponding Author:
Cecilia Cheng, PhD
Social and Health Psychology Laboratory
Department of Psychology
The University of Hong Kong
Centennial Campus, Pokfulam Road
Hong Kong
Hong Kong
Phone: 852 39174224
Email: ceci-cheng@hku.hk

Abstract

Background: COVID-19 has spread like wildfire across the globe, prompting many governments to impose unprecedented stay-at-home orders to limit its transmission. During an extended stay-at-home period, individuals may engage in more online leisure activities. Internet use is a double-edged sword that may have both desirable and undesirable effects on psychological well-being, and this study sought to disentangle adaptive from maladaptive internet use amidst this unusual health crisis.

Objective: The objectives of this study were to assess the prevalence of probable depression during the COVID-19 stay-at-home period and to test three hypothesized risk reduction or risk elevation mechanisms, namely social capital–accrual, escape-from-self, and time-displacement effects.

Methods: This study took place from March to May 2020 at the early stage of the pandemic. The study adopted a prospective design, with an online survey administered to 573 UK and 474 US adult residents at two assessment points 2 months apart.

Results: The prevalence of moderate to severe depression was 36% (bootstrap bias-corrected and accelerated [BCa] 95% CI 33%-39%) at Time 1 (ie, initial time point) and 27% (bootstrap BCa 95% CI 25%-30%) at Time 2 (ie, follow-up time point). The results supported the social capital–accrual hypothesis by showing that the approach coping style was inversely associated with Time 2 depression through its positive associations with both social networking and perceived family support. The results also supported the escape-from-self hypothesis by revealing that the avoidant coping style was positively associated with Time 2 depression through its positive associations with both gaming and cyberbullying victimization, but the serial mediation model was no longer significant after Time 1 depression and some demographic risk factors had been controlled for. Finally, the results supported the time-displacement hypothesis by showing that gaming was positively associated with Time 2 depression through its inverse associations with social networking and perceived family support.

Conclusions: During the extended stay-at-home period in the early stages of the COVID-19 pandemic, the prevalence of probable depression during the 2-month study period was high among the UK and US residents. Individuals with distinct coping styles may engage in different types of online leisure activities and perceive varying levels of social support, which are associated with risks of probable depression.

(J Med Internet Res 2020;22(12):e22740) doi:10.2196/22740

KEYWORDS

coping; coronavirus; COVID-19; cyberaggression; cybervictimization; epidemic; gaming; mental health; psychological well-being; social networking; social support

Introduction

Background Context

COVID-19 is an infectious disease caused by a novel strain of coronavirus known as SARS-CoV-2, which is deadly and highly transmissible [1,2]. Since the World Health Organization (WHO) reported the initial batch of confirmed COVID-19 cases on January 11, 2020, the disease has affected more than 68 million people and caused more than 1.6 million deaths in 191 countries and regions, globally, as of December 9, 2020 [3].

In response to the massive threat posed by COVID-19, numerous governments around the world have imposed a series of community-control measures in an attempt to curb its rapid transmission. For instance, the UK government implemented self-quarantine orders and school closures in March 2020 that lasted for around two months. During the same period, the majority of US states imposed similar stay-at-home orders. The residents of both countries were urged to avoid social gatherings and to stay at home, although they were allowed to go out for essential errands, such as buying groceries and attending medical appointments [4]. A US household survey revealed that the prevalence of probable depression during the early stage of the pandemic was 3 times higher than the prepandemic rate [5]. Systematic and meta-analytic reviews similarly showed that probable depression was prevalent (overall effect size of 34%, 95% CI 28%-41%) across the globe during the early stage of the pandemic [6,7] and that such alarming rates were 3-fold higher than the lifetime prevalence of depression from 1994 to 2014 (11%) [8]. Besides, sleep disturbance was also prevalent among the general public during the pandemic [9,10]. Apart from demographic factors such as gender and age, it is noteworthy that frequent social networking site (SNS) use was identified as a risk factor that heightened mental health problems [6,10].

Internet Use and Mental Health Issues

This study examined major online leisure activities that people engaged in during the stay-at-home period and, more importantly, the problems related to such activities. With rapid advances in information and communication technology and the emergence of affordable mobile devices (eg, smartphones and tablets), people are increasingly reliant on the internet for social networking, entertainment, information, and online purchases. Among the array of online leisure activities, social networking is the most popular, as reflected in the escalation of the number of SNS users over the past two decades [11]. Facebook, the most widely used SNS, has more than 2.7 billion active users worldwide as of August 2020 [12]. Another highly popular online leisure activity is gaming; there were more than 2.3 billion active video gamers all around the world in 2019 [13].

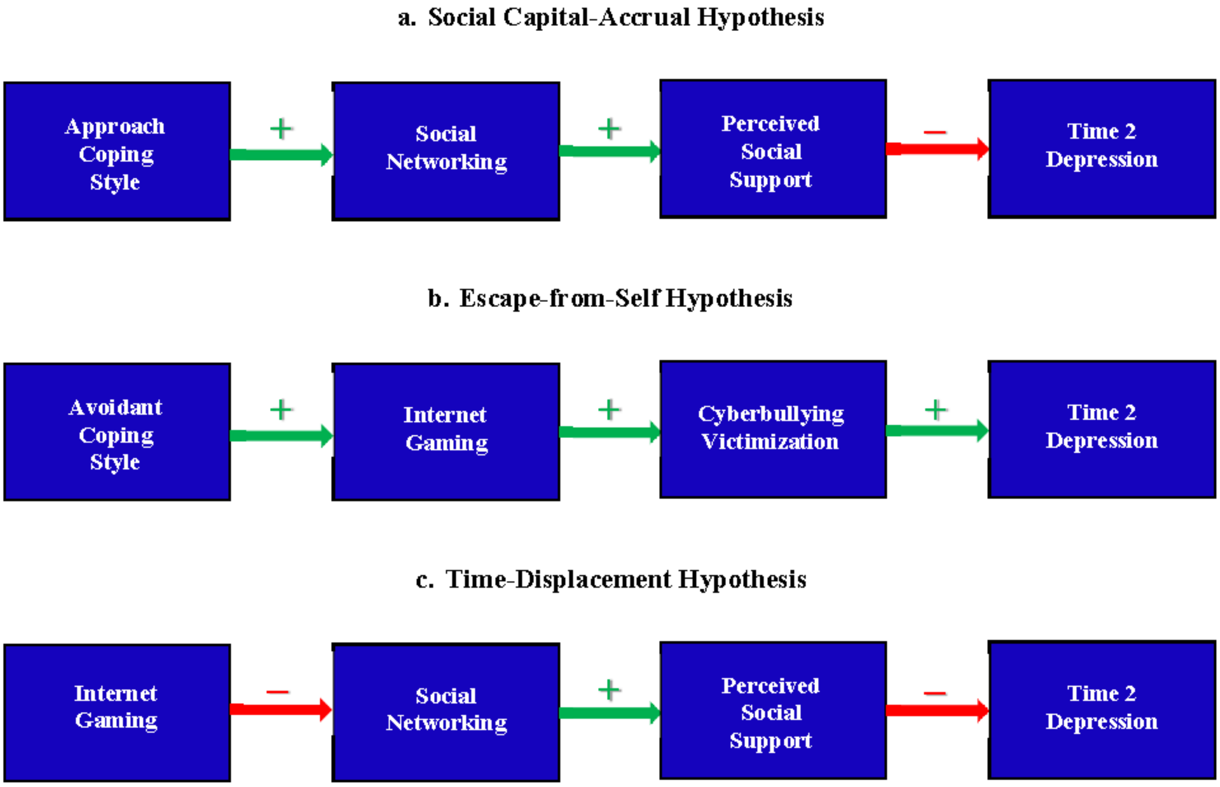
The use of the internet can be beneficial, allowing connections with other people without geographical or time constraints [14]. Support from social network members rendered online has been reported to alleviate depressive symptoms [15,16]. However, certain problems brought about by internet use also merit attention [17]. For instance, internet use may increase the likelihood of cybervictimization, such as impersonation and social exclusion, which in turn increases the risk of depression [18,19]. Moreover, excessive engagement in online games may disrupt daily functioning [20,21]. Internet use is, thus, a double-edged sword that can enhance and compromise mental health.

Risk Reduction and Elevation Mechanisms Underlying Internet Use

This study aimed at unveiling both risk reduction and risk elevation mechanisms underlying leisure-time internet use and depression during the COVID-19 pandemic. Figure 1 depicts the conceptual models that illustrate three hypothesized mechanisms. According to the conservation of resources theory [22,23], individuals strive to obtain and retain resources that facilitate goal attainment, and they feel distressed when such resources are lost or unavailable after making significant effort. Since the onset of the ongoing pandemic, the social resources of many people had been severely depleted while the stay-at-home orders were in place. As face-to-face interactions were not feasible, people attempted to regain their reduced social resources through the internet. SNS has emerged as a popular venue for rendering and gaining social support in the cyber era [15,24], and users' social support and subjective well-being were found to be increased after social networking [25,26].

A meta-analysis revealed considerable individual differences in the beneficial role of social networking in facilitating the accrual of online social capital [27]. Specifically, the meta-analytic findings showed that individuals who are motivated to use SNS to maintain contacts with members of their existing social networks tend to accrue more social capital that mitigates depressive symptoms. In a stressful encounter, an individual's coping style plays a pivotal role in influencing the display of cognitive and behavioral responses, which, in turn, have mental health implications [28,29]. In the coping literature, a broad conceptualization has been widely adopted to dichotomize these psychological responses into two general coping styles: approach versus avoidant [29,30]. Approach coping refers to a tendency to undertake direct actions as an attempt to confront or change a stressful event, whereas avoidant coping refers to a tendency to retreat or divert attention away from a stressful event [31,32]. As individuals characterized by an approach coping style tend to face the stressful encounter and take proactive actions to tackle problems [33,34], we predict that these individuals may actively utilize SNS to accrue more support from members of their social network and may, thus, be less vulnerable to depression (see Figure 1a).

Figure 1. Conceptual models summarizing three hypotheses tested in this study. Time 2 is the follow-up time point.



The meta-analytic findings further revealed that such mental health benefits, however, were absent among those who were motivated to use the internet as a refuge to avoid facing problems in the real world. According to the escaping-the-self theory [35], individuals tend to feel distressed when they realize that real-life outcomes are falling short of their expectations; they are motivated to evade such psychological discomfort by engaging in avoidant behaviors to divert their attention from existing problems. Their avoidant responses may elicit greater distress in the long run because the real-life problems remain intact [36,37]. Previous studies have consistently shown escapism as a major motivation of gaming that elicits adverse outcomes, such as cyberbullying victimization [38]. In the digital age, cyberbullying victimization is prevalent [39,40], and frequent such experiences enhance mental health risks [41,42]. In light of these theories and findings, we predict that individuals characterized by an avoidant coping style will tend to rely on gaming in an attempt to handle stress; however, their gaming behavior may increase their exposure to cyberbullying victimization, which may, in turn, enhance their susceptibility to depression (see Figure 1b).

It is noteworthy that investing more leisure time in gaming may reduce time spent social networking and, thus, the opportunities of reaping support from social network members. This notion stems from displacement theory [43], whose underlying assumption is that the total time spent on daily activities is constant. Hence, spending more time on a particular online activity results in less time on another. In light of the time-displacement theory, we predict that more time spent on gaming may be associated with less time spent on social networking and lower levels of social support, thus enhancing the risk of depression (see Figure 1c).

Study Setting and Aims

This study was conducted at an early stage of the COVID-19 pandemic, when stay-at-home orders were in effect in many countries, including the United Kingdom and the United States. The study aimed at assessing the prevalence of probable depression and testing three hypothesized mechanisms that would reduce or enhance risks of depression: social capital–accrual, escape-from-self, and time-displacement effects (see Figure 1).

Methods

Study Design and Settings

A prospective design with a quantitative approach was adopted in this study, which comprised two assessment points. Initial (Time 1) data collection took place from March 16 to 22, 2020, the week after the WHO declared the COVID-19 outbreak a global pandemic and stay-at-home orders were first implemented in various countries. Follow-up (Time 2) data collection then took place from May 18 to 24, 2020, during which stay-at-home orders were gradually lifted in some regions of the United Kingdom and the United States.

Study Sample and Eligibility Criteria

The participants were recruited through Prolific [44], a crowdsourcing platform for participant recruitment. Studies have shown that data generated via the crowdsourcing method were as reliable and valid as those obtained by offline methods, with participants reporting that they felt more comfortable sharing their personal data in crowdsourcing research than with other types of research [45,46]. Prolific was chosen because its participants were reported to have the most diverse demographic characteristics and to generate the highest data quality among popular crowdsourcing platforms [47,48].

Eligible participants were adults between the ages of 18 and 65 years who were living in the United Kingdom or the United States at the time of the study. Participants from other countries, those who did not take part in the follow-up assessment, and those who did not give informed consent were excluded. Data were collected from 1086 eligible participants, but 16 of them did not submit the survey, while 23 were timed out before they completed or submitted the survey (completion rate 96.4%). The final sample contained 1047 community adults.

Measures

Overview

The online survey contained a set of validated measures, all of which were chosen because they were short and designed for use in surveys administered in general populations. All of the measures were administered at Time 1; the measure assessing depression was administered at both time points. The measures were arranged in a randomized order to counterbalance potential order effects. Participants were prompted if there were missing responses.

Coping Style

Both approach and avoidant coping styles were measured by the Coping Strategies Inventory–Short Form [49]. Each of the coping subscales included four items, each of which was rated on a 5-point scale, ranging from 1 (*never*) to 5 (*almost always*). A higher composite score (Cronbach α=.79 for the approach coping style subscale and .89 for the avoidant coping style subscale) indicates the endorsement of a particular coping style.

Cyberbullying Victimization

Cyberbullying victimization was assessed by the cybervictimization subscale of the Cyber-Aggression and Cyber-Victimization Scale [50]. The subscale contained eight items, each of which was rated on a 5-point scale, ranging from 1 (*never*) to 5 (*very often*). A composite score (Cronbach α=.88 in this sample) was derived by aggregating the scores for all items, with a higher score representing a greater degree of cyberbullying victimization.

Perceived Social Support

Perceived family support and friend support were measured by the family and friend subscales of the Multidimensional Scale of Perceived Social Support [51]. Each subscale had four items, each of which was rated on a 7-point scale, ranging from 1 (*very strongly disagree*) to 7 (*very strongly agree*). A subscale score (Cronbach α=.94 for the family subscale and .95 for the friend subscale) was derived from the sum of all item scores, with a higher score representing the subjective appraisals of greater perceived support from a particular social group.

Depression

Each participant's level of depression was assessed by the Center for Epidemiological Studies–Depression Scale [52]. This depression scale comprised 20 items, each of which was rated on a 4-point scale, ranging from 0 (*rarely or none of the time*) to 3 (*all of the time*). A total depressive symptom score (Cronbach α=.84 and .80 at Time 1 and Time 2, respectively) was computed by summing all of the item scores, with a higher score representing a greater frequency and severity of depression. The conventional cutoff scores were 16 and over for mild depression and 23 and over for moderate to severe depression [52].

Time Allocated to Social Networking and Internet Gaming During the COVID-19 Pandemic

The participants reported the amount of leisure time (hours) they had dedicated on a typical day within the past week to two highly popular internet activities: social networking and gaming. Both of these items were found valid for assessing these online leisure activities [53,54].

Demographic Variables

At the end of the survey, the participants were asked to provide the following demographic information: age, sex, educational qualifications (ie, undergraduate degree holder or not), employment status (ie, employed, no need to work [ie, student, homemaker, or retired], or unemployed), marital status (ie, married or partnered or not), and ethnicity (ie, White or non-White). For conducting statistical analyses, the three-level employment status was recoded as a pair of dummy variables: employed (ie, 1=employed, 0=no need to work, 0=unemployed) and unemployed (ie, 0=employed, 0=no need to work, 1=unemployed). Age was recorded as a continuous variable, whereas all of the other demographic variables were dummy coded.

Study Procedures

Survey invitations were sent by Prolific to its members who met the demographic criteria. The survey was distributed through the Qualtrics survey system. Participation was entirely voluntary. Potential participants were told the survey aim and length, payment rate for completion (ie, £7.80 [US $10.50] per hour), and the ethical approval number and agent. They were also assured that all data collected would be anonymous and kept in strict confidentiality in the investigators' laboratory. All of the participants were requested to give informed consent prior to their participation at both time points, and were paid upon the completion of each survey. The data collected at the two time points were matched according to a unique code assigned to each participant by the survey platform.

Strategy of Analysis

All statistical analyses were performed with SPSS, version 26 (IBM Corp). Independent-samples *t* tests were conducted to examine whether there were any between-country differences (ie, United Kingdom vs United States) in the study variable. To test the three hypothesized mechanisms underlying the experience of Time 2 depression during the early stage of the COVID-19 pandemic, serial mediation analysis was performed using Model 6 of the SPSS macro PROCESS, version 3.5 [55], which used ordinary least squares procedures to estimate the hypothesized effects. Three sets of analyses were performed. First, initial tests for all three hypothesized serial mediation effects were conducted without any covariates. Second, the hypothesized effects were then tested with Time 1 depression as a covariate in order to capture temporal changes in depression level. Third, both Time 1 depression and relevant demographic variables were included as covariates to rule out any additional confounding factors. Recent reviews on depression assessed during the COVID-19 pandemic revealed some demographic characteristics (ie, age, gender, country, and finance-related factors) as risk factors that enhanced susceptibility to depression [5,6]. In addition, marital status was identified to be crucial for social connectedness [56]. Taken together, the following demographic risk factors were included as covariates in the third set of model testing: age; gender; country; employment status, which was represented by a pair of dummy variables (ie, employed or unemployed); and marital status.

In all of the mediation analyses, the scores on the predictor variables were centered to reduce multicollinearity [57]. Bootstrapping was conducted with 10,000 iterations (2-tailed significance), with the bootstrap bias-corrected and accelerated (BCa) interval employed to estimate the CIs. The hypothesized effects were deemed statistically significant if zero was excluded in the bootstrapped CIs.

Ethical Considerations

The research protocol was reviewed and approved by the human research ethics committee of the University of Hong Kong (approval No. EA2002033) prior to Time 1 data collection. All study procedures were carried out in accordance with the ethical rules of the Declaration of Helsinki of 1975, as revised in 2008.

Results

Sample Characteristics

The sample from this study consisted of 1047 adults residing in the United Kingdom and the United States at the time of the study. Participants from the two countries did not differ significantly in terms of the study variables (*P* values were above .14), except for Time 1 depression (t_{1043}=11.91, *P*<.001). The UK participants (mean score 22.32, SD 13.24) reported higher levels of depression at Time 1 than their US counterparts (mean score 18.11, SD 11.08). The two groups were pooled to enhance statistical power, but the demographic variable of country was included as a covariate in model testing.

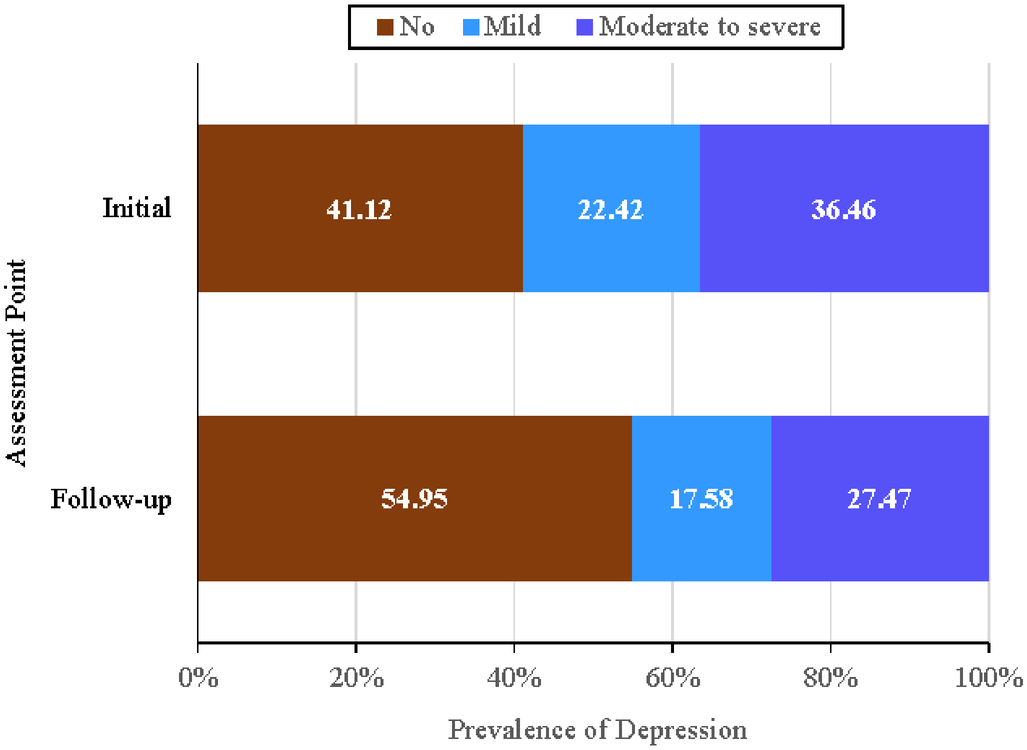
The pooled sample of 1047 participants comprised 481 men (45.9%) and 563 women (53.8%), as well as 3 participants (0.3%) who did not indicate their sex. The average age was 44.10 years (SD 12.59). Slightly more than half of the participants were undergraduate degree holders or above (596/1044, 57.1%) and were married or partnered (569/1036, 54.9%). Most participants were ethnically White (887/1041, 85.2%).

On average, the participants spent 1.35 hours (bootstrap BCa 95% CI 1.28-1.42) on social networking and 1.70 hours (bootstrap BCa 95% CI 1.59-1.81) on gaming.

Prevalence of Probable Depression at the Early Stage of the COVID-19 Pandemic

The average depression scores at the initial and follow-up time points were 20.41 (bootstrap BCa 95% CI 19.60-21.12) and 16.74 (bootstrap BCa 95% CI 16.06-17.42), respectively. Figure 2 illustrates the proportion of participants categorized as having no, mild, and moderate to severe depression. According to the standard cutoff scheme, the prevalence of moderate to severe depression was quite high: 36% (bootstrap BCa 95% CI 33%-39%) at Time 1 and 27% (bootstrap BCa 95% CI 25%-30%) at Time 2.

Figure 2. Prevalence of depression at the initial (March 16-22, 2020) and follow-up (May 18-24, 2020) assessment points during the COVID-19 pandemic.



Mediation Analyses Unveiling Risk Reduction and Elevation Mechanisms

The results of serial mediation analysis for testing the various hypotheses are summarized in Table 1. The various serial mediation models were first tested without any covariates entered (see the upper panel of Table 1). Consistent with the social capital–accrual hypothesis, the results showed that the indirect effects of the approach coping style on Time 2 depression via social networking and perceived social support—both family and peer support—were significant. In line with the escape-from-self hypothesis, the indirect effect of the avoidant coping style on Time 2 depression via gaming and cyberbullying was significant. Consistent with the time-displacement hypothesis, the indirect effects of gaming on Time 2 depression via social networking and perceived social support—both family and peer support—were significant.

A highly similar pattern of findings was obtained after controlling for Time 1 depression (see the middle panel of Table 1), except for the set of findings for testing the escape-from-self hypothesis. Specifically, this particular serial mediation model was no longer significant. We then performed post hoc tests for simple mediation effects (ie, PROCESS Model 4), and gaming was found to mediate the positive association between the avoidant coping style and Time 2 depression after the baseline depression level had been controlled for.

In the final set of models, Time 1 depression and some demographic risk factors of depression were included as covariates (see the lower panel of Table 1). Similar to the previous set of analyses, only the simple mediation effect of gaming between the avoidant coping style and Time 2 depression was significant. In addition, the two serial mediation effects with perceived peer support as a mediator became nonsignificant when the cluster of demographic risk factors had been entered. In summary, this final set of analyses showed strong support for both social capital–accrual and time-displacement hypotheses when perceived family support was a mediator.

Table 1. Summary of model testing for three hypotheses (N=1047).

Indirect effect	Effect	Bootstrap, SE	BCa 95% CI[a]	R^2
Models with no covariates				
H1[b]a: APC[c]→ SNW[d]→ PFS[e]→ Time 2 depression	–0.0144	0.0055	–0.0268 to –0.0055	0.2000
H1b: APC → SNW → PPS[f]→ Time 2 depression	–0.0171	0.0063	–0.0306 to –0.0063	0.2200
H2[g]: AVC[h]→ Gaming → CBV[i]→ Time 2 depression	0.0199	0.0084	0.0064 to 0.0391	0.2800
H3[j]a: Gaming → SNW → PFS → Time 2 depression	0.0431	0.0121	0.0227 to 0.0703	0.1700
H3b: Gaming → SNW → PPS → Time 2 depression	0.0340	0.0102	0.0164 to 0.0565	0.1700
Models with Time 1 depression as a covariate				
H1a: APC → SNW → PFS → Time 2 depression	–0.0129	0.0052	–0.0249 to –0.0046	0.2300
H1b: APC → SNW → PPS → Time 2 depression	–0.0152	0.0059	–0.0279 to –0.0051	0.2500
H2: AVC → Gaming → CBV → Time 2 depression[k]	0.0095	0.0065	–0.0003 to 0.0250	0.3000
H2 (post hoc test): AVC → Gaming → Time 2 depression	0.0296	0.0145	0.0054 to 0.0619	0.2200
H3a: Gaming → SNW → PFS → Time 2 depression	0.0375	0.0121	0.0174 to 0.0638	0.2000
H3b: Gaming → SNW → PPS → Time 2 depression	0.0306	0.0104	0.0131 to 0.0539	0.2000
Models with Time 1 depression and demographic risk factors as covariates[l]				
H1a: APC → SNW → PFS → Time 2 depression	–0.0037	0.0020	–0.0083 to –0.0008	0.5100
H1b: APC → SNW → PPS → Time 2 depression[k]	–0.0017	0.0014	–0.0051 to 0.0005	0.5000
H2: AVC → Gaming → CBV → Time 2 depression[k]	0.0053	0.0042	–0.0005 to 0.0156	0.5300
H2 (post hoc test): AVC → Gaming → Time 2 depression	0.0182	0.0104	0.0019 to 0.0417	0.5000
H3a: Gaming → SNW → PFS → Time 2 depression	0.0095	0.0044	0.0025 to 0.0197	0.5100
H3b: Gaming → SNW → PPS → Time 2 depression[k]	0.0042	0.0028	–0.0002 to 0.0107	0.5000

[a] Bias-corrected and accelerated (BCa) bootstrapped CIs, computed based on 10,000 bootstrap samples, were employed to interpret the significance of results for indirect effects instead of inferential tests [55].

[b] H1: social capital–accrual hypothesis, with either perceived family support (a) or perceived peer support (b).

[c] APC: approach coping style.

[d] SNW: social networking.

[e] PFS: perceived family support.

[f] PPS: perceived peer support.

[g] H2: escape-from-self hypothesis.

[h] AVC: avoidant coping style.

[i] CBV: cyberbullying victimization.

[j] H3: time-displacement hypothesis, with either perceived family support (a) or perceived peer support (b).

[k] This model resulted in nonsignificant findings.

[l] Demographic risk factor covariates include gender, age, country, employment status (ie, employed or unemployed), and marital status.

Discussion

Principal Findings

This study was conducted at a time when stay-at-home orders were in effect during the early stage of the COVID-19 pandemic. The findings indicate a high prevalence of probable depression immediately after the WHO's designation of the novel disease as a global pandemic. Specifically, more than one-third of our participants reported moderate to severe depression in the initial assessment period. Although the prevalence rate was lower at the second assessment that took place 2 months later, it remained at a relatively high level, with around one-quarter of the sample reporting moderate to severe depression at the follow-up period. These findings indicate that it was quite common for UK and US residents to experience some forms of depression when self-quarantining in their respective countries. The prevalence of probable depression obtained at both time points was about 3 times higher than the prepandemic prevalence [5]. Similarly, the prevalence of probable depression obtained at the initial time point was 3 times higher, and that obtained at the follow-up time point was about 2.5 times higher, than the lifetime prevalence of depression from 1994 to 2014 reported in a review [8].

Consistent with the social capital–accrual hypothesis, the findings show that individuals who report higher levels of approach coping tend to spend more leisure time on social networking and perceive a higher level of family support, which is related to lower subsequent levels of depression during the extended stay-at-home period. In contrast, individuals who report higher levels of avoidant coping tend to spend more time gaming, which is related to higher subsequent levels of depression. In line with the time-displacement hypothesis, gaming time is inversely associated with both social networking time and perceived family support, which is inversely associated with subsequent levels of depression.

Recommendations

These findings highlight the mental health implications of leisure-time internet use during the extended stay-at-home period at the early stage of the COVID-19 pandemic. Clinicians should be aware of individual differences in coping styles and assess whether the amount and type of internet use may increase or decrease risk for depression. Online professional support may be helpful to maintain psychological well-being if support from one's social network is unavailable or perceived to be scant during social isolation. If in-person appointments for mental health services cannot be made, the use of telemedicine can bridge this gap by strengthening adaptive coping with various types of stressors during the pandemic [58]. Further, the general public may benefit from increased awareness about mental health issues during the extended stay-at-home period and acquire practical strategies to reduce risk of depression, such as limiting sources of stress, social support seeking, and maintaining a regular routine [59].

As excessive leisure-time internet use, especially for gaming, can heighten risks of depression, clinicians are advised to expend more effort probing into the amount of leisure time their clients spend on the internet as part of a comprehensive mental health assessment, as well as to evaluate their clients' engagement in specific leisure activities (eg, social networking and gaming) that are likely to enhance or mitigate depression risk [60]. For instance, limiting the amount of time devoted to unconfirmed or questionable sources of COVID-19–related information, as well as maintaining a routine that does not involve excessive internet use, could reduce depressive symptoms [59].

The identification of mechanistic pathways leading to psychological issues during the pandemic has been proposed as a research priority owing to its relevance to refining interventions [61]. This study indicates that depression experienced during the extended stay-at-home period is a function of three crucial factors: coping style, the type of leisure-time internet use, and the type of social support. The three pathways identified in our study have practical implications. Firstly, the social capital–accrual pathway highlights the importance of the approach coping style and social networking in an attempt to elicit greater social support from family members. Secondly, the escape-from-self pathway specifies the maladaptive role of the avoidant coping style and gaming. Finally, the time-displacement pathway indicates that more time spent on gaming may reduce time spent on social networking and perceptions of family support. Clinicians who deliver psychological interventions to clients who lack family support should evaluate their clients' patterns of leisure-time internet use as a potential contributor to depression.

Study Limitations and Future Research Directions

Prior to concluding, several study limitations and directions for future research should be noted. First, the participants in this study were residents of the United Kingdom or the United States, both with some of the highest internet penetration rates in the world [62]. Accordingly, our findings may not be applicable to other COVID-19 hard-hit countries, particularly those with low internet penetration rates. Moreover, the participants were all residents of individualist countries, whose cultural values differ considerably from those shared by members of collectivist countries, such as Brazil and India [63-65]. Studies have shown considerable cultural differences in the motivation of social networking [66] and in the prevalence of internet addiction and internet gaming disorder [67,68]. Although the prevalence rates of probable depression of our two samples were comparable to those reported in a systematic review during the COVID-19 pandemic in eight countries across four continents [6], future studies should expand the scope of our research by recruiting participants from an array of countries with varying levels of internet penetration and cultural values to allow for more extensive multinational and cross-cultural comparisons.

Second, the prospective study was conducted during an early stage of the COVID-19 pandemic when unprecedented stay-at-home orders were in place. It is noteworthy that the findings reflect the participants' initial psychological responses to the pandemic and may, thus, not be generalizable to other waves of the pandemic or to the postpandemic period. Previous studies have shown social isolation for extended periods to have long-term undesirable psychological impacts [69,70], and mental health issues fluctuated drastically across various waves of a disease outbreak [71]. Follow-up studies with a longer time horizon should be conducted to examine the potential chronic impact once the current drastic control measures have been lifted.

Finally, our study focused on the two most popular online leisure activities, namely social networking and gaming, because validated measures are only available for assessing time spent on these two types of activities. Nowadays, individuals also browse the internet for other purposes (eg, watching videos and movies and shopping), and it is worthwhile to expand the scope of internet use if validated measures are available for tapping these activities as well. As interpersonal interactions are minimal in many of these alternative activities, the pattern of findings may differ from those yielded in this study. Also, studies have demonstrated that interpersonal behaviors tend to differ among players of distinct game genres (eg, single-player vs multiplayer and cooperative vs competitive gamers) [72-75], and the pattern of findings may vary according to the type of games played on a regular basis. Future research may benefit from a nuanced analysis that takes the game genre and type of internet activity into consideration to enhance explanatory and predictive utility.

Conclusions

During the extended stay-at-home period mandated to deal with COVID-19 transmission, depression was found to be prevalent. The findings reported herein indicate that leisure-time internet use can reduce or increase risks of depression, depending largely on the coping style of the users. As increasing numbers of people worldwide are being required to stay at home, additional studies should be conducted to obtain a more nuanced picture of the impacts of coping style, engagement in online leisure activities, and family dynamics on mental health.

Acknowledgments

This work was supported by the Hong Kong Research Grants Council's General Research Fund (grant No. 17400714) and the University of Hong Kong's Seed Fund for Basic Research (grant No. 201711159216). We thank Yvonne Tsui and Ellick Wong for statistical advice and comments on an earlier draft. Thanks also go to Sylvia Lam, Janice Leung, and Olivia Qiu for research and clerical assistance.

Conflicts of Interest

None declared.

References

1. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020 Apr;26(4):450-452 [FREE Full text] [doi: 10.1038/s41591-020-0820-9] [Medline: 32284615]
2. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020 Mar;55(3):105924 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105924] [Medline: 32081636]
3. Global map. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-12-09]
4. Gajanan M. What does 'shelter in place' mean? Here's what life is like under the mandate. Time. 2020 Mar 20. URL: https://time.com/5806477/what-is-shelter-in-place/ [accessed 2020-03-25]
5. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
6. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
7. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. Global Health 2020 Jul 06;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
8. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of depression in the community from 30 countries between 1994 and 2014. Sci Rep 2018 Feb 12;8(1):2861 [FREE Full text] [doi: 10.1038/s41598-018-21243-x] [Medline: 29434331]
9. Cheng C, Ebrahimi OV, Lau Y. Maladaptive coping with the infodemic and sleep disturbance in the COVID-19 pandemic. J Sleep Res 2020 Nov 27:e13235. [doi: 10.1111/jsr.13235] [Medline: 33247519]
10. Cellini N, Canale N, Mioni G, Costa S. Changes in sleep pattern, sense of time and digital media use during COVID-19 lockdown in Italy. J Sleep Res 2020 Aug;29(4):e13074 [FREE Full text] [doi: 10.1111/jsr.13074] [Medline: 32410272]
11. Clement J. Number of social network users worldwide from 2017 to 2025. Statista. 2020 Nov 24. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2020-12-08]
12. Clement J. Facebook: Number of monthly active users worldwide 2008-2020. Statista. 2020 Nov 24. URL: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/ [accessed 2020-12-08]
13. Global Esports Market Report. 5th edition. Amsterdam, the Netherlands: Newzoo; 2019. URL: https://newzoo.com/insights/trend-reports/newzoo-global-esports-market-report-2019-light-version/ [accessed 2020-09-27]
14. Seabrook EM, Kern ML, Rickard NS. Social networking sites, depression, and anxiety: A systematic review. JMIR Ment Health 2016 Nov 23;3(4):e50 [FREE Full text] [doi: 10.2196/mental.5842] [Medline: 27881357]
15. Chan MS, Cheng C. Explaining personality and contextual differences in beneficial role of online versus offline social support: A moderated mediation model. Comput Human Behav 2016 Oct;63:747-756. [doi: 10.1016/j.chb.2016.05.058]
16. Billedo C, Kerkhof P, Finkenauer C, Ganzeboom H. Facebook and face-to-face: Examining the short- and long-term reciprocal effects of interactions, perceived social support, and depression among international students. J Comput Mediat Commun 2019;24(2):73-89. [doi: 10.1093/jcmc/zmy025]
17. Sigerson L, Li AY, Cheung MW, Cheng C. Examining common information technology addictions and their relationships with non-technology-related addictions. Comput Human Behav 2017 Oct;75:520-526. [doi: 10.1016/j.chb.2017.05.041]

18. Fossati P, Hinfray S, Fall A, Lemogne C, Rotge J. Emotional and cognitive consequences of social rejection: An entry door to major depression. In: Baune BT, Harmer C, editors. Cognitive Dimensions of Major Depressive Disorder. Oxford, UK: Oxford University Press; 2019:211-223.
19. Liu C, Liu Z, Yuan G. The longitudinal influence of cyberbullying victimization on depression and posttraumatic stress symptoms: The mediation role of rumination. Arch Psychiatr Nurs 2020 Aug;34(4):206-210. [doi: 10.1016/j.apnu.2020.05.002] [Medline: 32828350]
20. Wang H, Cheng C. New perspectives on the prevalence and associated factors of gaming disorder in Hong Kong community adults: A generational approach. Comput Human Behav 2021 Jan;114:106574. [doi: 10.1016/j.chb.2020.106574]
21. Mills DJ, Allen JJ. Self-determination theory, internet gaming disorder, and the mediating role of self-control. Comput Human Behav 2020 Apr;105:106209. [doi: 10.1016/j.chb.2019.106209]
22. Hobfoll SE. Conservation of resources theory: Its implication for stress, health, and resilience. In: Folkman S, editor. The Oxford Handbook of Stress, Health, and Coping. New York, NY: Oxford University Press; 2011:127-147.
23. Le BM, Impett EA, Kogan A, Webster GD, Cheng C. The personal and interpersonal rewards of communal orientation. J Soc Pers Relat 2012 Dec 12;30(6):694-710. [doi: 10.1177/0265407512466227]
24. Wang G, Zhang W, Zeng R. WeChat use intensity and social support: The moderating effect of motivators for WeChat use. Comput Human Behav 2019 Feb;91:244-251. [doi: 10.1016/j.chb.2018.10.010]
25. Oh HJ, Ozkaya E, LaRose R. How does online social networking enhance life satisfaction? The relationships among online supportive interaction, affect, perceived social support, sense of community, and life satisfaction. Comput Human Behav 2014 Jan;30:69-78. [doi: 10.1016/j.chb.2013.07.053]
26. Li L, Peng W. Transitioning through social media: International students' SNS use, perceived social support, and acculturative stress. Comput Human Behav 2019 Sep;98:69-79. [doi: 10.1016/j.chb.2019.03.011]
27. Cheng C, Wang H, Sigerson L, Chau C. Do the socially rich get richer? A nuanced perspective on social network site use and online social capital accrual. Psychol Bull 2019 Jul;145(7):734-764. [doi: 10.1037/bul0000198] [Medline: 31094537]
28. Cheng C, Lau HB, Chan MS. Coping flexibility and psychological adjustment to stressful life changes: A meta-analytic review. Psychol Bull 2014 Nov;140(6):1582-1607. [doi: 10.1037/a0037913] [Medline: 25222637]
29. Roth S, Cohen LJ. Approach, avoidance, and coping with stress. Am Psychol 1986 Jul;41(7):813-819. [doi: 10.1037//0003-066x.41.7.813] [Medline: 3740641]
30. Cheng C, Chau C. When to approach and when to avoid? Functional flexibility is the key. Psychol Inq 2019 Sep 24;30(3):125-129. [doi: 10.1080/1047840x.2019.1646040]
31. Cheng C. Cognitive and motivational processes underlying coping flexibility: A dual-process model. J Pers Soc Psychol 2003 Feb;84(2):425-438. [Medline: 12585814]
32. Cheng C. Dialectical thinking and coping flexibility: A multimethod approach. J Pers 2009 Apr;77(2):471-494. [doi: 10.1111/j.1467-6494.2008.00555.x] [Medline: 19220723]
33. Folkman S, Lazarus R, Dunkel-Schetter C, DeLongis A, Gruen R. The dynamics of a stressful encounter. In: Higgins ET, Kruglanski AW, editors. Motivational Science: Social and Personality Perspectives. Philadelphia, PA: Psychology Press; 2000:111-127.
34. Cheng C, Hui W, Lam S. Psychosocial factors and perceived severity of functional dyspeptic symptoms: A psychosocial interactionist model. Psychosom Med 2004;66(1):85-91. [doi: 10.1097/01.psy.0000106885.40753.c1] [Medline: 14747642]
35. Baumeister R. Escaping the Self: Alcoholism, Spirituality, Masochism, and Other Flights from the Burden of Selfhood. New York, NY: Basic Books; 1991.
36. Cheng C, Sun P, Mak K. Internet addiction and psychosocial maladjustment: Avoidant coping and coping inflexibility as psychological mechanisms. Cyberpsychol Behav Soc Netw 2015 Sep;18(9):539-546. [doi: 10.1089/cyber.2015.0121] [Medline: 26348815]
37. Gan Y, Zhang Y, Wang X, Wang S, Shen X. The coping flexibility of neurasthenia and depressive patients. Pers Individ Dif 2006 Apr;40(5):859-871. [doi: 10.1016/j.paid.2005.09.006]
38. McInroy LB, Mishna F. Cyberbullying on online gaming platforms for children and youth. Child Adolesc Soc Work J 2017 Apr 19;34(6):597-607. [doi: 10.1007/s10560-017-0498-0]
39. Chang F, Chiu C, Miao N, Chen P, Lee C, Huang T, et al. Online gaming and risks predict cyberbullying perpetration and victimization in adolescents. Int J Public Health 2015 Feb;60(2):257-266. [doi: 10.1007/s00038-014-0643-x] [Medline: 25549613]
40. Fryling M, Cotler J, Rivituso J, Mathews L, Pratico S. Cyberbullying or normal game play? Impact of age, gender, and experience on cyberbullying in multi-player online gaming environments: Perceptions from one gaming forum. J Inf Systs Appl Res 2015 Apr;8(1):4-18 ISSN: 1946-1836 [FREE Full text]
41. Bowditch L, Chapman J, Naweed A. Do coping strategies moderate the relationship between escapism and negative gaming outcomes in World of Warcraft (MMORPG) players? Comput Human Behav 2018 Sep;86:69-76. [doi: 10.1016/j.chb.2018.04.030]
42. Wang M, Yogeeswaran K, Andrews NP, Hawi DR, Sibley CG. How common is cyberbullying among adults? Exploring gender, ethnic, and age differences in the prevalence of cyberbullying. Cyberpsychol Behav Soc Netw 2019 Nov;22(11):736-741. [doi: 10.1089/cyber.2019.0146] [Medline: 31697598]

43. Nie N, Hillygus D. The impact of internet use on sociability: Time-diary findings. IT and Society. 2002. URL: https://sites.duke.edu/hillygus/files/2014/05/v01i01a01.pdf [accessed 2020-03-27]
44. Prolific. URL: https://www.prolific.co/ [accessed 2020-03-02]
45. Sescleifer AM, Francoisse CA, Lin AY. Systematic review: Online crowdsourcing to assess perceptual speech outcomes. J Surg Res 2018 Dec;232:351-364. [doi: 10.1016/j.jss.2018.06.032] [Medline: 30463741]
46. Strickland JC, Victor GA. Leveraging crowdsourcing methods to collect qualitative data in addiction science: Narratives of non-medical prescription opioid, heroin, and fentanyl use. Int J Drug Policy 2020 Jan;75:102587. [doi: 10.1016/j.drugpo.2019.10.013] [Medline: 31751813]
47. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
48. Palan S, Schitter C. Prolific.ac—A subject pool for online experiments. J Behav Exp Finance 2018 Mar;17:22-27. [doi: 10.1016/j.jbef.2017.12.004]
49. Tobin DL, Holroyd KA, Reynolds RV, Wigal JK. The hierarchical factor structure of the coping strategies inventory. Cogn Ther Res 1989 Aug;13(4):343-361. [doi: 10.1007/BF01173478]
50. Shapka JD, Maghsoudi R. Examining the validity and reliability of the cyber-aggression and cyber-victimization scale. Comput Human Behav 2017 Apr;69:10-17. [doi: 10.1016/j.chb.2016.12.015]
51. Zimet GD, Dahlem NW, Zimet SG, Farley GK. The Multidimensional Scale of Perceived Social Support. J Pers Assess 1988 Mar;52(1):30-41. [doi: 10.1207/s15327752jpa5201_2]
52. Radloff LS. The CES-D Scale. Appl Psychol Meas 1977 Jun 1;1(3):385-401. [doi: 10.1177/014662167700100306]
53. Sigerson L, Cheng C. Scales for measuring user engagement with social network sites: A systematic review of psychometric properties. Comput Human Behav 2018 Jun;83:87-105. [doi: 10.1016/j.chb.2018.01.023]
54. Sigerson L, Li AY, Cheung MW, Luk JW, Cheng C. Psychometric properties of the Chinese Internet Gaming Disorder Scale. Addict Behav 2017 Nov;74:20-26 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.031] [Medline: 28558336]
55. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. 2nd Edition. New York, NY: The Guilford Press; 2018.
56. Holt-Lunstad J, Robles TF, Sbarra DA. Advancing social connection as a public health priority in the United States. Am Psychol 2017 Sep;72(6):517-530 [FREE Full text] [doi: 10.1037/amp0000103] [Medline: 28880099]
57. Aiken LS, West SG. Multiple Regression: Testing and Interpreting Interactions. Newbury Park, CA: SAGE Publications; 1991.
58. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The role of telehealth in reducing the mental health burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
59. Fiorillo A, Gorwood P. The consequences of the COVID-19 pandemic on mental health and implications for clinical practice. Eur Psychiatry 2020 Apr 01;63(1):e32 [FREE Full text] [doi: 10.1192/j.eurpsy.2020.35] [Medline: 32234102]
60. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. Lancet Psychiatry 2020 Apr;7(4):300-302 [FREE Full text] [doi: 10.1016/S2215-0366(20)30073-0] [Medline: 32085840]
61. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
62. World internet usage and population statistics. Internet World Stats. 2020 May 31. URL: https://www.internetworldstats.com/stats.htm [accessed 2020-06-02]
63. Cheng C, Jose PE, Sheldon KM, Singelis TM, Cheung MWL, Tiliouine H, et al. Sociocultural differences in self-construal and subjective well-being: A test of four cultural models. J Cross Cult Psychol 2010 Sep 09;42(5):832-855. [doi: 10.1177/0022022110381117]
64. Cheng C, Cheung MW, Montasem A, 44 members of the International Network of Well-Being Studies. Explaining differences in subjective well-being across 33 nations using multilevel models: Universal personality, cultural relativity, and national income. J Pers 2016 Feb;84(1):46-58. [doi: 10.1111/jopy.12136] [Medline: 25234240]
65. Triandis HC. Individualism-collectivism and personality. J Pers 2001 Dec;69(6):907-924. [doi: 10.1111/1467-6494.696169] [Medline: 11767823]
66. Sariyska R, Lachmann B, Cheng C, Gnisci A, Sergi I, Pace A, et al. The motivation for Facebook use – Is it a matter of bonding or control over others? J Individ Differ 2019 Jan 01;40(1):26-35. [doi: 10.1027/1614-0001/a000273]
67. Cheng C, Cheung MW, Wang H. Multinational comparison of internet gaming disorder and psychosocial problems versus well-being: Meta-analysis of 20 countries. Comput Human Behav 2018 Nov;88:153-167. [doi: 10.1016/j.chb.2018.06.033]
68. Cheng C, Li AY. Internet addiction prevalence and quality of (real) life: A meta-analysis of 31 nations across seven world regions. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):755-760 [FREE Full text] [doi: 10.1089/cyber.2014.0317] [Medline: 25489876]
69. Adams KB, Sanders S, Auth EA. Loneliness and depression in independent living retirement communities: Risk and resilience factors. Aging Ment Health 2004 Nov;8(6):475-485. [doi: 10.1080/13607860410001725054] [Medline: 15724829]

70. Golden J, Conroy RM, Bruce I, Denihan A, Greene E, Kirby M, et al. Loneliness, social support networks, mood and wellbeing in community-dwelling elderly. Int J Geriatr Psychiatry 2009 Jul;24(7):694-700. [doi: 10.1002/gps.2181] [Medline: 19274642]
71. Cheng C, Cheung MWL. Psychological responses to outbreak of severe acute respiratory syndrome: A prospective, multiple time-point study. J Pers 2005 Feb;73(1):261-285 [FREE Full text] [doi: 10.1111/j.1467-6494.2004.00310.x] [Medline: 15660679]
72. Sung H, Sigerson L, Cheng C. Social capital accumulation in location-based mobile game playing: A multiple-process approach. Cyberpsychol Behav Soc Netw 2017 Aug;20(8):486-493. [doi: 10.1089/cyber.2017.0222] [Medline: 28806121]
73. Lisón JF, Cebolla A, Guixeres J, Álvarez-Pitti J, Escobar P, Bruñó A, et al. Competitive active video games: Physiological and psychological responses in children and adolescents. Paediatr Child Health 2015 Oct;20(7):373-376 [FREE Full text] [doi: 10.1093/pch/20.7.373] [Medline: 26526217]
74. O'Connor EL, Longman H, White KM, Obst PL. Sense of community, social identity and social support among players of massively multiplayer online games (MMOGS): A qualitative analysis. J Community Appl Soc Psychol 2015 Jan 15;25(6):459-473. [doi: 10.1002/casp.2224]
75. Greitemeyer T, Cox C. There's no “I” in team: Effects of cooperative video games on cooperative behavior. Eur J Soc Psychol 2013 Apr 10;43(3):224-228. [doi: 10.1002/ejsp.1940]

Abbreviations

BCa: bias-corrected and accelerated
SNS: social networking site
WHO: World Health Organization

Edited by G Eysenbach; submitted 22.07.20; peer-reviewed by T Cahill, R Ho; comments to author 19.09.20; revised version received 08.10.20; accepted 14.12.20; published 24.12.20.

Please cite as:
Cheng C, Lau YC, Luk JW
Social Capital–Accrual, Escape-From-Self, and Time-Displacement Effects of Internet Use During the COVID-19 Stay-at-Home Period: Prospective, Quantitative Survey Study
J Med Internet Res 2020;22(12):e22740
URL: http://www.jmir.org/2020/12/e22740/
doi:10.2196/22740
PMID:33320824

Original Paper

Anxiety and Suicidal Thoughts During the COVID-19 Pandemic: Cross-Country Comparative Study Among Indonesian, Taiwanese, and Thai University Students

Iqbal Pramukti[1,2*], PhD; Carol Strong[3*], MPH, PhD; Yajai Sitthimongkol[4], PhD; Agus Setiawan[5], PhD; Moses Glorino Rumambo Pandin[6], Psi, MSi, MPhil, MPsi, DPhil; Cheng-Fang Yen[7], MD, PhD; Chung-Ying Lin[8], PhD; Mark D Griffiths[9], PhD; Nai-Ying Ko[2,10], PhD

[1]Faculty of Nursing, Universitas Padjadjaran, West Java, Indonesia
[2]International Doctoral Program in Nursing, Department of Nursing, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[3]Department of Public Health, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[4]Faculty of Nursing, Mahidol University, Bangkok, Thailand
[5]Faculty of Nursing, Universitas Indonesia, Jakarta, Indonesia
[6]Faculty of Humanities, Universitas Airlangga, East Java, Indonesia
[7]Department of Psychiatry, Kaohsiung Medical University Hospital & School of Medicine College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
[8]Institute of Allied Health Sciences, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[9]International Gaming Research Unit, Psychology Department, Nottingham Trent University, Nottingham, United Kingdom
[10]Department of Nursing, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[*]these authors contributed equally

Corresponding Author:
Chung-Ying Lin, PhD
Institute of Allied Health Sciences, National Cheng Kung University Hospital
College of Medicine
National Cheng Kung University
1 University Road
Tainan, 701
Taiwan
Phone: 886 6 2353535 ext 5720
Email: cylin36933@gmail.com

Abstract

Background: The COVID-19 pandemic has negatively affected the mental health of university students.

Objective: This study examined the psychological responses toward COVID-19 among university students from 3 countries—Indonesia, Taiwan, and Thailand.

Methods: We used a web-based, cross-sectional survey to recruit 1985 university students from 5 public universities (2 in Indonesia, 1 in Thailand, and 1 in Taiwan) via popular social media platforms such as Facebook, LINE, WhatsApp, and broadcast. All students (n=938 in Indonesia, n=734 in Thailand, and n=313 in Taiwan) answered questions concerning their anxiety, suicidal thoughts (or sadness), confidence in pandemic control, risk perception of susceptibility to infection, perceived support, resources for fighting infection, and sources of information in the context of the COVID-19 pandemic.

Results: Among the 3 student groups, Thai students had the highest levels of anxiety but the lowest levels of confidence in pandemic control and available resources for fighting COVID-19. Factors associated with higher anxiety differed across countries. Less perceived satisfactory support was associated with more suicidal thoughts among Indonesian students. On the other hand, Taiwanese students were more negatively affected by information gathered from the internet and from medical staff than were Indonesian or Thai students.

Conclusions: Our findings suggest that health care providers in Thailand may need to pay special attention to Thai university students given that high levels of anxiety were observed in this study population. In addition, health care providers should establish a good support system for university students, as the results of this study indicate a negative association between support and suicidal thoughts.

(J Med Internet Res 2020;22(12):e24487) doi:10.2196/24487

KEYWORDS
anxiety; COVID-19; cross-country; suicidal thoughts; university students

Introduction

The COVID-19 outbreak was declared as a pandemic on March 11, 2020, by the World Health Organization [1]. It has affected over 100 countries worldwide, including many countries in the Asia-Pacific region. As of October 20, 2020, nearly 40 million confirmed cases of COVID-19 were reported globally, with over 9 million cases reported in the Asia-Pacific region [1]. In Indonesia, Thailand, and Taiwan, in particular, the total number of confirmed COVID-19 cases was over 137,000; 3300; and 480, respectively, as of August 16, 2020 [1-3].

With increasing numbers of COVID-19 infections and associated deaths, psychological responses toward COVID-19 have become an important topic for health care providers [4,5]. Indeed, studies on different populations (including pregnant women and their husbands, general population, and children) have reported increased psychological problems during the COVID-19 pandemic [6-8]. Moreover, instruments assessing psychological responses specifically designed in relation to COVID-19 have been developed and validated [9-14]. Therefore, the importance of assessing psychological responses toward COVID-19 and their associated factors has been highlighted in the extant literature. More specifically, adverse effects of elevated psychological distress may trigger reassurance-seeking and compulsive-checking behaviors in response to potential threats of COVID-19 infection, which, in turn, may have an impact on the daily lives of individuals who impede preventive behaviors [15,16].

Although university students worldwide are affected by the COVID-19 pandemic, their psychological responses may be different because of cultural differences and varying levels of the COVID-19 crisis across countries. For example, Taiwan undertook early action to prevent COVID-19 transmission and maintained an almost normal living state without having to implement a national lockdown [10,15,17]. Such a situation may have mitigated negative psychological responses among Taiwanese university students. On the other hand, COVID-19–related fatality rate in Indonesia was found to be relatively high (8.9%), and it may have more negatively affected university students than other countries [18].

This study examined university students' psychological responses toward COVID-19 for the following reasons. First, the COVID-19 pandemic has become a major life stressor that has direct and indirect influences on students' psychological well-being. The direct influences include the students' emotional feelings toward COVID-19 (eg, fear of being infected) [19-21], and the indirect influences include the government's reaction in relation to transmission control (eg, feeling depressed when receiving threatening COVID-19 news or feeling lonely because of social distancing) [7]. Second, university students are at a critical life-transition period (ie, school-to-work transition [22]). Most of them are expected to begin their careers after graduation by applying the skills they have learned at the university [23]. However, the COVID-19 pandemic may interfere with their plans and, subsequently, affect their future careers. Therefore, university students need support in maintaining good mental health conditions in order to deal with the COVID-19 pandemic, as well as avoid any negative consequences later in life.

Therefore, in this study, we applied a combination of the health belief model (HBM) [24] and protection motivation theory (PMT) [25], to examine potential independent variables for explaining psychological responses among university students. The HBM posits that perceived susceptibility, perceived severity, and perceived benefits are the major contributors enabling individuals to take specific health behavioral actions. Moreover, the PMT posits that individuals' health-related behaviors are triggered by their psychological distress (eg, anxiety and fear). Therefore, the factors proposed by the HBM are potential independent variables that help explain the psychological responses of individuals. In this study, we assessed the factors proposed by the HBM as follows: perceived susceptibility and perceived severity were assessed using risk perception of susceptibility to COVID-19; perceived benefits were assessed using confidence in pandemic control (ie, controlling the pandemic is a benefit for the individual to fight COVID-19), perceived support (ie, having support is a benefit for the individual to fight COVID-19), and perceived sufficiency of resources (ie, having sufficient resources is a benefit for the individual to fight COVID-19).

In addition to applying the HBM [24] and PMT [25], the existing literature on COVID-19 also suggests that these independent variables could contribute to an individual's psychological responses. For instance, a higher confidence in fighting COVID-19 may be associated with more adaptive psychological responses when dealing with COVID-19 [17,26]. Fear of COVID-19 was found to be lower when the country had better control of COVID-19 cases. As people are known to have fear of COVID-19 [27] and being stigmatized if they are infected [28], they may have more negative psychological responses when they perceive higher risk of having COVID-19. In addition, previous studies have made recommendations to provide sufficient resources such as personal protective equipment (PPE) and support to assist individuals in combating psychological problems during the COVID-19 outbreak [29,30].

Apart from the HBM and PMT, the literature proposes the importance of obtaining accurate COVID-19 information from trusted sources. Indeed, misinformation concerning COVID-19 is associated with a greater fear of the disease [15,16]. Hou et al [31] reported that rumors and misinformation shared on the internet may induce panic behaviors such as purchasing herbal remedies via web-based shopping and storing them. Based on the HBM, PMT, and the existing literature, we hypothesized that confidence in COVID-19 pandemic control, risk perception of susceptibility to COVID-19, perceived satisfactory support, perceived sufficiency of resources for fighting COVID-19, and sources of obtaining COVID-19 information are all potential factors that may be associated with psychological responses to COVID-19.

To the best of our knowledge, there have been few cross-country comparisons concerning psychological responses among university students. In order address this gap in the literature, we compared the psychological responses toward COVID-19 and its related factors among university students in 3 different countries—Indonesia, Thailand, and Taiwan. These 3 countries were selected for the comparative study because they are all East Asian countries that share similar cultures and values (ie, Confucianism) [32]. In addition, a previous study on country variations concerning depression symptoms found similar prevalence of low self-confidence in these 3 countries [33]. However, the 3 countries had different outcomes and policies in minimizing the impact of the COVID-19 pandemic. Therefore, it would be useful to apply the same model to 3 countries that have similar cultures and values but had implemented different measures in combating COVID-19.

The primary outcomes of this study were different psychological responses such as anxiety and suicidal thoughts. Moreover, we examined other related factors (ie, confidence in COVID-19 pandemic control, risk perception of susceptibility to COVID-19, perceived satisfactory support, perceived sufficiency of resources for fighting COVID-19, and sources of obtaining COVID-19 information) to understand their associations with psychological responses among different university student groups from different countries.

Methods

Study Design, Participants, and Data Collection

A multicountry, web-based cross-sectional study was conducted in 5 public universities. The sample comprised students in Indonesia (2 universities), Taiwan (1 university), and Thailand (1 university). Participants were recruited through popular social media platforms operational in these countries, including Facebook, LINE, WhatsApp, and Broadcast, from April 10 to June 30, 2020. Only participants aged 20 years and above were eligible for this study. Before beginning the survey, the participants were asked to log in with their personal email addresses in order to avoid having participants repeat the survey more than once.

We obtained approvals from the Research Ethics Committee in each of the 3 countries studied (ie, University of Indonesia [SK-139/UN2.F12.D1.2.1/ETIK 2020] for Indonesia, National Cheng Kung University Hospital [A-EX-109-019] for Taiwan, and Mahidol University [COA No. MU-COVID 2020.006/1205] for Thailand). Thereafter, data collection was initiated. Participation in the survey was voluntary, and survey responses were anonymously collected. Study participants were given no incentive for participation. Participants who agreed to participate in the study completed the web-based survey in their native languages: Bahasa (Indonesian students), Mandarin (Taiwanese students), and Thai (Thai students).

Measures

Outcome Variables: Anxiety and Suicidal Thoughts (or Sadness)

The State-Trait Anxiety Inventory (STAI) was used to assess the anxiety levels of the participants toward the COVID-19 pandemic. The STAI comprises 10 items rated on a 4-point Likert scale ("not at all," "a little," "somewhat," and "very much"). A lower score on the STAI indicates a lower level of anxiety [34]. Suicidal thoughts in the past week were assessed for Taiwanese and Indonesian students on a 5-point Likert scale ("not at all," "mild," "moderate," "severe," and "very severe"). A lower score on this scale indicates a lower level of suicidal ideation. Similarly, sadness experienced in the past week was assessed for Thai students on the same 5-point Likert scale. Although sadness is not a direct concept of suicidal thoughts, it can be viewed as a proxy of suicidal thoughts for Thai students. More specifically, we considered that when a Thai student feels sad for a prolonged period, their physiological and psychological functions will be disturbed and may further lead to depressive symptoms with a high risk of suicidal ideation.

Independent Variables

Confidence in Pandemic Control

Confidence in pandemic control included confidence concerning personal, city, and the university in handling the pandemic situation. This variable was rated on a 5-point Likert scale from 1 ("not confident at all") to 5 ("very confident") [35]. The Cronbach alpha of the 3 items concerning confidence in pandemic control indicated very good internal reliability (Cronbach α=.83).

Risk Perception of Susceptibility to COVID-19

Risk perception of susceptibility to COVID-19 included both absolute susceptibility and relative susceptibility, both of which were rated on a 7-point Likert scale from 1 ("not at all susceptible") to 7 ("very susceptible") [36]. The Cronbach alpha value of the 2 items concerning risk perception of susceptibility to COVID-19 indicated very good internal reliability (Cronbach α=.80).

Perceived Support

Perceived support assesses support received from families, classmates, and faculties. The 3 items were rated on a 5-point Likert scale from 1 ("not satisfied at all") to 5 ("very satisfied") [37]. The Cronbach alpha value of the 3 items concerning perceived support indicated acceptable internal reliability (Cronbach α=.69).

Perceived Sufficiency of Resources for Fighting COVID-19

Perceived sufficiency of resources included perceived sufficiency of PPE, COVID-19 information, money, medical resources, and psychological support. All items were rated on a 3-point Likert scale from 1 ("insufficient") to 3 ("sufficient") [38]. The Cronbach alpha value of the 5 items concerning perceived sufficiency of resources for fight COVID-19 indicated very good internal reliability (Cronbach α=.76).

Sources of COVID-19 Information

Seeking information from various sources included information from the internet, medical staff, and family [39]. All items were rated on a 3-point Likert scale from 1 (“never”) to 3 (“always”).

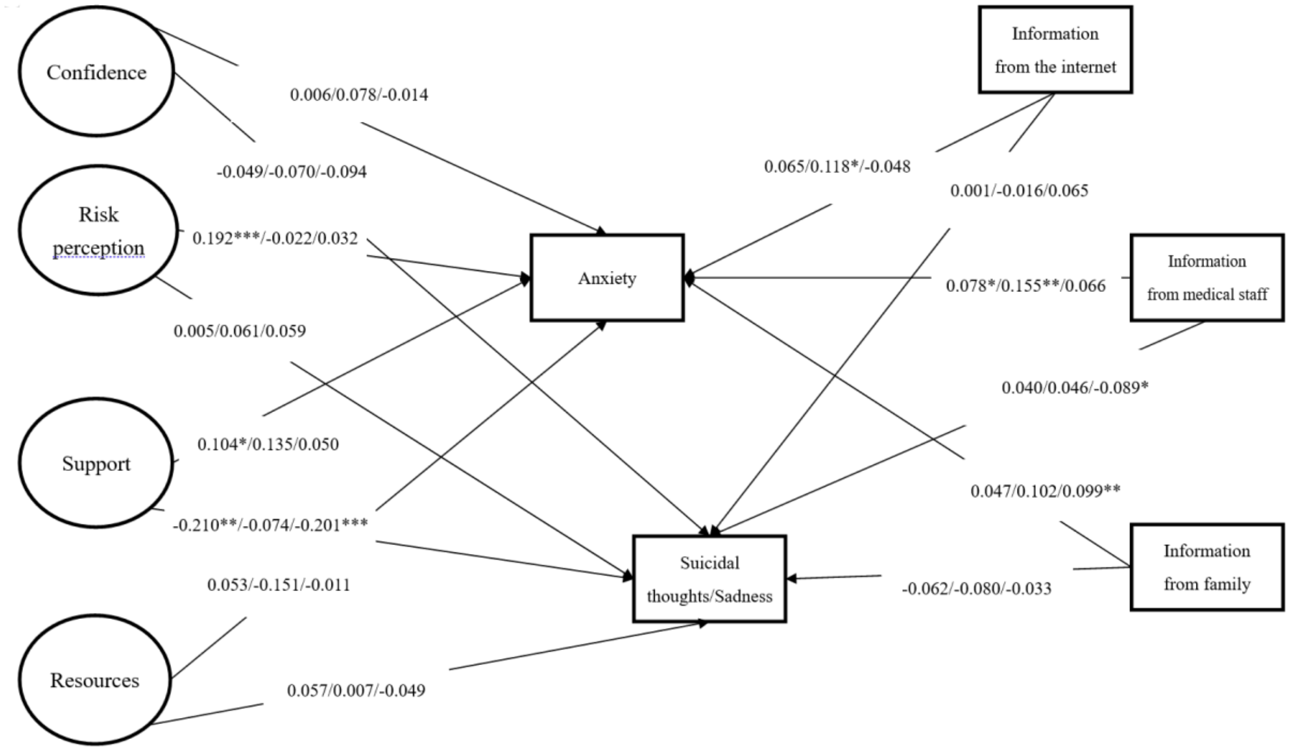
Demographic Information

Sociodemographic data such as gender, age, and education level were collected from the survey responses. Students with a Bachelor’s qualification were classified as undergraduates, and those with a Master’s qualification or above were classified as postgraduates. Students were also asked whether their major subjects of study were health-related or not.

Statistical Analysis

Analyses of variance (ANOVA) and χ^2 tests were used to examine the differences concerning dependent variables, independent variables, and controlled variables. Posthoc comparisons with Bonferroni adjustments were used when an overall statistical significance was observed in ANOVAs or χ^2 tests. Pearson correlations were then used to understand the bivariate correlations between the studied variables for all participants. Multigroup structural equation modeling (SEM) was then applied to examine how the independent variables explained anxiety and suicidal thoughts (for Taiwanese and Indonesian students) or anxiety and sadness (for Thai students; see Figure 1). In the multigroup SEM, confidence in fighting COVID-19, risk perception of susceptibility to COVID-19, perceived satisfactory support, and perceived sufficiency of resources were latent variables, whereas anxiety, suicidal thoughts (or sadness), sources of COVID-19 information (internet, medical staff, and family), and gender were manifest variables. Weighted least squares mean and variance adjusted estimator was used in the multigroup SEM to address the nature of Likert-type scales used in the study measures. Fit indices such as comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) were used to determine whether the multigroup SEM is supported (CFI and TLI should be 0.9 or above; RMSEA and SRMR should 0.08 or less [40,41]).

Figure 1. Results of the proposed model showing standardized coefficients among Indonesian, Taiwanese, and Thai students. **P*<.05, ***P*<.01, ****P*<.001. Gender was controlled for in the model. Suicidal thoughts were assessed for Taiwanese and Indonesian students; sadness was assessed for Thai students. Confidence was constructed using (1) confidence to deal with the pandemic, (2) confidence of the city to deal with the pandemic, and (3) confidence of students to deal with the pandemic. Risk perception was constructed using (1) perceived absolute susceptibility to COVID-19 and (2) perceived relative susceptibility to COVID-19. Support was constructed using (1) satisfaction with family support, (2) satisfaction with friend support, and (3) satisfaction with university support. Resource was constructed using (1) sufficiency of personal protective equipment, (2) sufficiency of information, (3) sufficiency of money, (4) sufficiency of medical resources, and (5) sufficiency of psychological support. Fit indices: χ2378=1035.12; *P*<.001; comparative fit index=0.92; Tucker-Lewis index=0.90; root mean square error of approximation=0.051 (90% CI 0.048-0.055); standardized root mean square residual index=0.053.



Results

Among the 1985 university students, 938 (47.2%) were Indonesian, 734 (37%) were Thai, and 313 (15.8%) were Taiwanese. Approximately 80% (1588/1985) of all study participants were female, with 82.2% of the participants majoring in medical-related programs (1631/1985) and 80.9% studying at the undergraduate level (1605/1985). The compositions of these aforementioned demographics were significantly different among the 3 groups. More specifically, the Taiwanese student sample comprised more males, the Indonesian student sample comprised more medical students, and the Thai student sample comprised more postgraduates (Table 1).

Table 1. Participant characteristics in different groups (N=1985).

Characteristic	Value			*F* test (*df1*, *df2*)	Chi-square (*df*)	*P* value	Post-hoc comparison
	Indonesian students (n=938)	Taiwanese students (n=313)	Thai students (n=734)				
Gender, male, n (%)	134 (14.3)	117 (37.4)	146 (19.9)	N/A[a]	78.76 (*2*)	<.001	2>3>1
Medical student, yes, n (%)	841 (89.7)	233 (74.4)	557 (75.9)	N/A	565.21 (*2*)	<.001	1>2,3
Postgraduate, yes, n (%)	125 (13.3)	37 (11.8)	218 (29.7)	N/A	92.65 (*2*)	<.001	3>1,2
Anxiety, mean (SD)	2.33 (0.48)	2.08 (0.42)	2.55 (0.43)	129.19 (*2, 1982*)	N/A	<.001	3>1>2
Confidence[b], mean (SD)	2.75 (0.79)	2.37 (0.68)	2.01 (0.86)	172.43 (*2, 1982*)	N/A	<.001	1>2>3
Perceived risk[c], mean (SD)	3.35 (1.16)	3.47 (0.96)	3.15 (1.16)	10.95 (*2, 1982*)	N/A	<.001	1,2>3
Support[d], mean (SD)	4.20 (0.59)	3.89 (0.68)	3.48 (0.70)	259.10 (*2, 1982*)	N/A	<.001	1>2>3
Resources[e], mean (SD)	1.39 (0.50)	1.80 (0.31)	1.31 (0.49)	122.37 (*2, 1982*)	N/A	<.001	2>1>3
Internet[f], mean (SD)	2.73 (0.49)	2.70 (0.49)	2.82 (0.42)	10.59 (*2, 1982*)	N/A	<.001	3>1,2
Medical staff[g], mean (SD)	2.63 (0.58)	2.34 (0.67)	2.37 (0.68)	1.55 (*2, 1982*)	N/A	.21	--
Family[h], mean (SD)	2.52 (0.60)	2.18 (0.59)	2.22 (0.61)	66.52 (*2, 1982*)	N/A	<.001	1>2,3

[a] N/A: not applicable.

[b] Confidence: confidence in pandemic control.

[c] Perceived risk: risk perception of susceptibility to COVID-19.

[d] Support: perceived satisfactory support from family, friends, or university.

[e] Resources: sufficiency of resources.

[f] Internet: COVID-19 information received from the internet.

[g] Medical staff: COVID-19 information received from medical staff.

[h] Family: COVID-19 information received from family.

The differences in independent and outcome variables across the 3 student groups are also shown in Table 1. Our results showed that Thai students had the highest levels of anxiety, the lowest levels of confidence in fighting COVID-19, and the least sufficient resources among the 3 student groups. On the other hand, Indonesian students had the highest levels of risk perception of susceptibility to COVID-19 and perceived satisfactory support from different sources among the 3 student groups. Moreover, compared to the other groups, Thai students received more COVID-19 information from the internet, and

Indonesian students received more information from medical staff and family (Table 1).

Correlations between independent variables, outcome variables, and controlled variables are presented in Table 2. We found that anxiety was significantly associated with confidence in pandemic control (r=−.08; *P*<.001); risk perception of susceptibility to COVID-19 (r=.07; *P*=.003); sufficiency of resources (r=.08; *P*<.001); and receiving information from the internet (r=.10; *P*<.001), medical staff (r=.11; *P*<.001), and family (r=.08; *P*<.001). Suicidal thoughts were significantly associated with confidence in pandemic control (r=-.28; *P*<.001), perceived satisfactory support (r=-.36; *P*<.001), sufficient resources (r=-.21; *P*<.001), and receiving information from the internet (r=.06; *P*=.006) and family (r=-.12; *P*<.001).

Multigroup SEM showed satisfactory fit indices (CFI=0.92; TLI=0.90; RMSEA=0.051; SRMR=0.053). Regarding the path coefficients for Indonesian students, higher risk perception of susceptibility to COVID-19, greater perceived satisfactory support, and receiving more information from medical staff significantly explained higher levels of anxiety. Less perceived satisfactory support was the only independent variable that significantly explained more frequent suicidal thoughts among the Indonesian students.

Regarding the path coefficients for Taiwanese students, receiving more information from the internet and medical staff significantly explained the higher levels of anxiety observed in this study group. No independent variables significantly explained suicidal thoughts among Taiwanese students.

Finally, regarding the path coefficients for Thai students, only receiving more information from family significantly explained the higher levels of anxiety observed in this study group. Less perceived satisfactory support and receiving less information from medical staff significantly explained the more frequent sadness reported by Thai students (Figure 1).

Table 2. Correlation matrix (Pearson r and 2-tailed *P* values) for studied variables (N=1985).

Variable	Gender	Anxiety	Confidence	Perceived risk	Support	Resources	Internet	Medical staff	Family
Gender									
r	1	−.08	0.09	−.05	−.07	0.04	−.03	−.02	−.12
P value	—[a]	<.001	<.001	.02)*	.001	-0.06	-0.17	-0.5	<.001
Anxiety		—							
r	−.08	1	−.08	0.07	−.03	−.08	0.1	0.11	0.08
P value	<.001	—	<.001	.003	-0.17	<.001	<.001	<.001	<.001
Confidence[b]									
r	0.09	−.08	1	0.22	−.35	0.25	0.01	−.10	−.19
P value	<.001	<.001	—	<.001	<.001	<.001	-0.8	<.001	<.001
Perceived risk[c]									
r	−.05	0.07	0.22	1	−.01	−.13	−.06	0.1	−.08
P value	.02)*	.003	<.001	—	−0.74	<.001	.008	<.001	<.001)**
Support[d]									
r	−.07	−.03	−.35	−.01	1	0.22	0.04	0.09	0.27
P value	.001	-0.17	<.001	-0.74	—	<.001	-0.06	<.001	<.001
Resources[e]									
r	0.04	−.08	0.25	−.13	0.22	1	0.01	−.08	−.10
P value	-0.06	<.001	<.001	<.001	<.001	—	-0.65	.001	<.001
Internet[f]									
r	−.03	0.1	0.01	−.06	0.04	0.01	1	0.14	0.27
P value	-0.17	<.001	-0.8	.008	-0.06	-0.65	—	<.001	<.001
Medical staff[g]									
r	−.02	0.11	−.10	0.1	0.09	−.08	0.14	1	0.22
P value	-0.5	<.001	<.001	<.001	<.001	.001	<.001	—	<.001
Family[h]									
r	−.12	0.08	−.19	−.08	0.27	−.10	0.27	0.22	1
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	—

[a] Not applicable.

[b] Confidence: confidence in pandemic control.

[c] Perceived risk: risk perception of susceptibility to COVID-19.

[d] Support: perceived satisfactory support from family, friends, or university.

[e] Resources: sufficiency of resources.

[f] Internet: COVID-19 information received from the internet.

[g] Medical staff: COVID-19 information received from medical staff.

[h] Family: COVID-19 information received from family.

Discussion

Principal Findings

This study showed that, among the 3 student groups compared, Thai university students had the greatest negative psychological responses (ie, the highest level of anxiety), whereas Taiwanese students had the lowest negative psychological responses. Confidence in pandemic control, sufficiency of resources, and receiving COVID-19 information from the internet and family were all factors associated with both anxiety and suicidal thoughts in the overall study population. Moreover, factors associated with higher levels of psychological responses considerably differed across the 3 countries. For example, less perceived satisfactory support was associated with more suicidal thoughts among Indonesian students and more sadness among Thai students.

Thai students had the highest levels of anxiety among the 3 study groups, which may be attributed to the low confidence they expressed in pandemic control and the lack of resources for fighting COVID-19. Indeed, the correlation results indicated that anxiety and suicidal thoughts were negatively associated with confidence in pandemic control and sufficiency of resources. The main reason for the lowest levels of anxiety and suicidal thoughts among Taiwanese students could be the early reaction by the Taiwan government to control COVID-19 infection [17,42], which substantially decreased disease transmission rate. Consequently, the effects of COVID-19 on daily life were less substantial in Taiwan than in Indonesia and Thailand.

When comparing the various sources of COVID-19 information in relation to anxiety, we found that Taiwanese students were highly affected by COVID-19 information gathered from the internet and from medical staff. This finding was similar to that of another study on the general population in Taiwan, which found that receiving COVID-19 information from the internet was associated with poorer psychological well-being [43,44]. Previous research has indicated that the more an individual gathers internet-based COVID-19 information, the higher is the impact on the individual's anxiety level, a phenomenon termed "cyberchondria" [43,44]. Seeking health information on the internet has been the most common method to obtain health information even before the COVID-19 pandemic [45]. However, the content and quality of health information available on the internet regarding COVID-19 can vary depending on the region. For example, frequencies of recommendations regarding COVID-19 prevention, such as "wash your hands frequently" or "stay home if you feel unwell," appearing on the internet were significantly different between Spain and the United States [46].

The proposed model in this study was partially supported because confidence in pandemic control and sufficiency of resources were both associated with more positive psychological responses to the pandemic. This finding indicates that perceived benefits in the HBM is important for university students in overcoming their psychological challenges during the COVID-19 pandemic period. Moreover, the benefits concerning COVID-19 control (ie, confidence in COVID-19 control and resource sufficiency) appeared to be more important than the benefits of others' emotional support (ie, perceived support). Moreover, perceived satisfactory support appeared to be an important factor in our study given the contrasting findings concerning anxiety and suicidal thoughts in some cultures. Lower perceived satisfactory support was associated with greater suicidal thoughts among Indonesian students and more sadness among Thai students, but more perceived support was associated with higher anxiety among Indonesian students. Perceived support may prevent individuals from having suicidal thoughts or sadness, but it may result in increased anxiety due to sharing of COVID-19 information repeatedly in a smaller social network. Moreover, information on the internet may even contain misinformation concerning COVID-19 [47]. The general public needs to know where on the internet to seek accurate information related to COVID-19 and not to constantly check for new information, as this could negatively affect their psychological health. However, individuals are still encouraged to seek information because accurate information can help them to engage in appropriate preventive behaviors. Therefore, we recommend that the public should seek information moderately. In some cultures, such as among the Thai university students in this study, receiving more COVID-19 information from medical staff was associated with less sadness. This finding suggests that a more reliable source of COVID-19 information may subsequently reduce suicidal thoughts among Thai students, given that feeling sad or experiencing a low mood is one of the depressive symptoms that could lead to suicide. Future studies should also analyze the content, frequency, and the various types of COVID-19 information available online that may be directly associated with anxiety. This knowledge could be used to promote additional sources on COVID-19 information for the general population such that they can find reliable and accurate information.

The amount of information may be amplified due to echo chamber effects, wherein information is disseminated among those who know each other very well, and individuals believe and trust in the information shared because they are very familiar with each other [48]. Thus, if misinformation was spread and exacerbated in their networks, it is likely that their anxiety levels will increase. However, if individuals can obtain accurate COVID-19 information through their close social network, it could lead to a reduction in suicidal thoughts. The findings of our study also highlight that the risks of university students having suicidal thoughts during the COVID-19 pandemic should not be overlooked.

Based on the findings of this study, there are several implications. First, these findings suggest there is a need for health care providers to help improve the psychological health of university students by providing them with reliable information to increase their confidence in COVID-19 pandemic control and provide sufficient resources in dealing with the resulting psychological impact. More specifically, psychoeducation to assist university students in understanding how the city and university are controlling and minimizing the spread of COVID-19 may be one method for improving students' confidence in COVID-19 control. Additionally, programs highlighting preventive COVID-19 behaviors may also increase students' confidence in controlling COVID-19 at an individual level. Second, health care providers need to provide sufficient PPE, accurate COVID-19 information, monetary and medical resources, and psychological support to students. This would help students perceive the benefits of these resources in minimizing the impact of COVID-19 and decreasing their psychological distress. Third, health care providers should encourage students to obtain COVID-19 information from reputable and trusted sources, such as from medical staff, rather than from the internet and social media, as misinformation on the internet may be difficult to identify and may have a negative impact on the students' psychological health. In addition, health care providers also need to be proactive in correcting misinformation spread among university students.

Study Limitations

There are some limitations to this study. First, the majority of the university students in the present study were female majoring in medical-related programs. Therefore, the generalizability of the findings is restricted. Future studies with university students majoring in other programs (eg, social science and engineering) are therefore needed. Moreover, the present findings are arguably biased because, in general, medical professionals are subject to experiencing more psychological impacts due to the nature of their work requiring them to be at the frontline, especially during the ongoing COVID-19 pandemic. Second, although the present study recruited participants from different East Asian countries, the sample representativeness is restricted due to the sample sizes being different in the 3 countries studied. More specifically, the Taiwanese sample was relatively smaller (15.8%) than the other two samples. Therefore, the country comparisons are limited. Third, all the measures were assessed using self-reports received via a web-based survey. The biases of social desirability and common method variance may therefore influence these results. Fourth, confidence in pandemic control was assessed by combining 3 different domains (ie, personal, university, and city). Given that confidence in oneself can considerably differ from that in how a university or city handles a public health emergency situation, the confidence in pandemic control examined in the present study was not specific. Therefore, future studies should consider separating the confidence in different domains and examining the effects of each type of confidence. Finally, the response scales used in the present study were different for different variables (eg, the items concerning confidence in pandemic control used a 5-point scale, and items concerning risk perception of susceptibility to COVID-19 used a 7-point scale). The use of different response scales may have affected instrument reliability and validity. However, this may not be a serious problem because instruments using different response scales within the same measure (eg, Short-Form 36) have been shown to have good psychometric properties [49].

Conclusions

In conclusion, this study demonstrated that university students from different countries had different levels of psychological responses relating to the COVID-19 pandemic. Moreover, receiving more COVID-19 information appears to increase the anxiety levels among university students, but not in relation to suicidal thoughts. Receiving less satisfactory support was found to be associated with more suicidal thoughts among Indonesian students and greater sadness among Thai students. Therefore, health care providers need to establish a good support system for university students to get through the current pandemic.

Acknowledgments

We acknowledge financial support received from the Ministry of Science and Technology (MOST109-2327-B-006-005) and National Health Research Institute, Taiwan (NHRI-EX108-10625EI), and Southeast and South Asia and Taiwan Universities Joint Research Scheme (NCKU 33).

Authors' Contributions

IP, CS, C-FY, C-YL, and N-YK conceptualized the study. IP, CS, C-YL, and N-YK wrote the first draft. IP, YS, AS, MP, C-FY, and N-YK designed the online survey and collected data. C-YL analyzed the data. CS, C-YL, MDG, and N-YK interpreted the data. IP, CS, YS, AS, MP, C-FY, MDG, and N-YK critically reviewed the manuscript. MDG was responsible for the final editing. IP and CS contributed equally to the study; N-YK and C-YL contributed equally to the study. All the authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Coronavirus (COVID-19) outbreak. World Health Organization - Western Pacific. URL: https://www.who.int/westernpacific/emergencies/covid-19 [accessed 2020-12-15]
2. WHO. Coronavirus disease (COVID-19) Situation Report – 209. World Health Organization. 2020 Aug 16. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200816-covid-19-sitrep-209.pdf?sfvrsn=5dde1ca2_2 [accessed 2020-12-15]
3. Taiwan - coronavirus cases. Worldometer. 2020. URL: https://www.worldometers.info/coronavirus/country/taiwan/ [accessed 2020-12-15]
4. Lin C. Social reaction toward the 2019 novel coronavirus (COVID-19). Soc Health Behav 2020;3(1):1. [doi: 10.4103/shb.shb_11_20]
5. Maaravi Y, Heller B. Not all worries were created equal: the case of COVID-19 anxiety. Public Health 2020 Aug;185:243-245 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.032] [Medline: 32688099]
6. Ahorsu DK, Imani V, Lin C, Timpka T, Broström A, Updegraff JA, et al. Associations between fear of COVID-19, mental health, and preventive behaviours across pregnant women and husbands: An actor-partner interdependence modelling. Int J Ment Health Addict 2020 Jun 11:1-15 [FREE Full text] [doi: 10.1007/s11469-020-00340-x] [Medline: 32837427]

7. Chen I, Chen C, Pakpour AH, Griffiths MD, Lin C. Internet-related behaviors and psychological distress among schoolchildren during COVID-19 school suspension. J Am Acad Child Adolesc Psychiatry 2020 Oct;59(10):1099-1102.e1. [doi: 10.1016/j.jaac.2020.06.007] [Medline: 32615153]
8. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus diseases (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
9. Ahorsu DK, Lin C, Imani V, Saffari M, Griffiths MD, Pakpour AH. The fear of COVID-19 scale: development and initial validation. Int J Ment Health Addiction 2020 Mar 27. [doi: 10.1007/s11469-020-00270-8]
10. Chang K, Hou W, Pakpour AH, Lin C, Griffiths MD. Psychometric testing of three COVID-19-related scales among people with mental illness. Int J Ment Health Addict 2020 Jul 11:1-13 [FREE Full text] [doi: 10.1007/s11469-020-00361-6] [Medline: 32837442]
11. Pakpour AH, Griffiths MD, Chang K, Chen Y, Kuo Y, Lin C. Assessing the fear of COVID-19 among different populations: A response to Ransing et al. (2020). Brain Behav Immun 2020 Oct;89:524-525 [FREE Full text] [doi: 10.1016/j.bbi.2020.06.006] [Medline: 32512132]
12. Lee SA. Coronavirus anxiety scale: a brief mental health screener for COVID-19 related anxiety. Death Stud 2020;44(7):393-401. [doi: 10.1080/07481187.2020.1748481] [Medline: 32299304]
13. Lee SA. How much "Thinking" about COVID-19 is clinically dysfunctional? Brain Behav Immun 2020 Jul;87:97-98 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.067] [Medline: 32353520]
14. Taylor S, Landry CA, Paluszek MM, Fergus TA, McKay D, Asmundson GJ. Development and initial validation of the COVID Stress Scales. J Anxiety Disord 2020 May;72:102232 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102232] [Medline: 32408047]
15. Chang K, Strong C, Pakpour AH, Griffiths MD, Lin C. Factors related to preventive COVID-19 infection behaviors among people with mental illness. J Formos Med Assoc 2020 Dec;119(12):1772-1780 [FREE Full text] [doi: 10.1016/j.jfma.2020.07.032] [Medline: 32773260]
16. Lin C, Broström A, Griffiths MD, Pakpour AH. Investigating mediated effects of fear of COVID-19 and COVID-19 misunderstanding in the association between problematic social media use, psychological distress, and insomnia. Internet Interv 2020 Sep;21:100345 [FREE Full text] [doi: 10.1016/j.invent.2020.100345] [Medline: 32868992]
17. Lin M, Cheng Y. Policy actions to alleviate psychosocial impacts of COVID-19 pandemic: Experiences from Taiwan. Soc Health Behav 2020;3(2):72. [doi: 10.4103/shb.shb_18_20]
18. Setiati S, Azwar M. COVID-19 and Indonesia. Acta Med Indones 2020 Jan;52(1):84-89 [FREE Full text] [Medline: 32291377]
19. Akdeniz G, Kavakci M, Gozugok M, Yalcinkaya S, Kucukay A, Sahutogullari B. A survey of attitudes, anxiety Status, and protective behaviors of the university students during the COVID-19 outbreak in Turkey. Front Psychiatry 2020;11:695 [FREE Full text] [doi: 10.3389/fpsyt.2020.00695] [Medline: 32760303]
20. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
21. Zhai Y, Du X. Addressing collegiate mental health amid COVID-19 pandemic. Psychiatry Res 2020 Jun;288:113003 [FREE Full text] [doi: 10.1016/j.psychres.2020.113003] [Medline: 32315885]
22. Koen J, Klehe U, Van Vianen AE. Training career adaptability to facilitate a successful school-to-work transition. Journal of Vocational Behavior 2012 Dec;81(3):395-408. [doi: 10.1016/j.jvb.2012.10.003]
23. Kim S, Ahn T, Fouad N. Family influence on Korean students' career decisions. Journal of Career Assessment 2015 Aug 18;24(3):513-526. [doi: 10.1177/1069072715599403]
24. Green E, Murphy E. Health belief model. In: Cockerham WC, Dingwall R, Quah S, editors. The Wiley Blackwell encyclopedia of health, illness, behavior, and society. Hoboken, NJ: John Wiley & Sons; 2014.
25. Rogers RW. A protection motivation theory of fear appeals and attitude change. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
26. Li D, Ko N, Chen Y, Wang P, Chang Y, Yen C. Confidence in coping with COVID-19 and its related factors among the public in Taiwan. Psychiatry Clin Neurosci 2020 Nov;74(11):608-610 [FREE Full text] [doi: 10.1111/pcn.13128] [Medline: 32786034]
27. Mamun MA, Griffiths MD. First COVID-19 suicide case in Bangladesh due to fear of COVID-19 and xenophobia: Possible suicide prevention strategies. Asian J Psychiatr 2020 Jun;51:102073 [FREE Full text] [doi: 10.1016/j.ajp.2020.102073] [Medline: 32278889]
28. Thakur V, Jain A. COVID 2019-suicides: A global psychological pandemic. Brain Behav Immun 2020 Aug;88:952-953 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.062] [Medline: 32335196]
29. Greenberg N, Docherty M, Gnanapragasam S, Wessely S. Managing mental health challenges faced by healthcare workers during covid-19 pandemic. BMJ 2020 Mar 26:m1211. [doi: 10.1136/bmj.m1211]
30. Shrivastava S, Shrivastava P. COVID-19 pandemic: Responding to the challenge of global shortage of personal protective equipment. Soc Health Behav 2020;3(2):70. [doi: 10.4103/shb.shb_17_20]

31. Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, et al. Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: infodemiology study. J Med Internet Res 2020 Aug 03;22(8):e21143 [FREE Full text] [doi: 10.2196/21143] [Medline: 32701460]
32. Munkh-Ulzii MM, Van KP. Empirical models of herding behaviour for Asian countries with Confucian culture. In: Kreinovich V, Sriboonchitta S, Chakpitak N, editors. Predictive Econometrics and Big Data. TES 2018. Studies in Computational Intelligence. vol 753. Cham: Springer; 2018.
33. Chee K, Tripathi A, Avasthi A, Chong M, Xiang Y, Sim K, et al. Country variations in depressive symptoms profile in Asian countries: Findings of the Research on Asia Psychotropic Prescription (REAP) studies. Asia Pac Psychiatry 2015 Sep;7(3):276-285. [doi: 10.1111/appy.12170] [Medline: 25641910]
34. Chung S, Long C. A study of the revised State-Trait Anxiety Inventory. Psychol Testing 1984;31(1):27-36.
35. Voeten H. Standard questionnaire on risk perception of an infectious disease outbreak. In: Effective communication in outbreak management for Europe. The Netherlands: Municipal Public Health Service Rotterdam-Rijnmond (GGD); Nov 2015.
36. Liao Q, Cowling B, Lam W, Ng DMW, Fielding R. Anxiety, worry and cognitive risk estimate in relation to protective behaviors during the 2009 influenza A/H1N1 pandemic in Hong Kong: ten cross-sectional surveys. BMC Infect Dis 2014 Mar 27;14:169 [FREE Full text] [doi: 10.1186/1471-2334-14-169] [Medline: 24674239]
37. Li D, Ko N, Chen Y, Wang P, Chang Y, Yen C, et al. COVID-19-related factors associated with sleep disturbance and suicidal thoughts among the Taiwanese public: a Facebook survey. Int J Environ Res Public Health 2020 Jun 22;17(12):4479 [FREE Full text] [doi: 10.3390/ijerph17124479] [Medline: 32580433]
38. Glanz K, Rimer B, Viswanath K, editors. Health Behavior: Theory, Research, and Practice. 5th Edition. San Francisco, CA: Jossey-Bass; 2015.
39. Wang P, Lu W, Ko N, Chen YL, Li DJ, Chang YP, et al. COVID-19-related information sources and the relationship with confidence in people coping with COVID-19: Facebook survey study in Taiwan. J Med Internet Res 2020 Jun 05;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]
40. Lin Y, Fung XCC, Tsai M, Strong C, Hsieh Y, Lin C. Insufficient physical activity and overweight: does caregiver screen-viewing matter? J Child Fam Stud 2018 Sep 15;28(1):286-297. [doi: 10.1007/s10826-018-1247-5]
41. Yam C, Pakpour AH, Griffiths MD, Yau W, Lo CM, Ng JMT, et al. Psychometric testing of three Chinese online-related addictive behavior instruments among Hong Kong university students. Psychiatr Q 2019 Mar;90(1):117-128. [doi: 10.1007/s11126-018-9610-7] [Medline: 30328020]
42. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing. JAMA 2020 Apr 14;323(14):1341-1342. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
43. Muse K, McManus F, Leung C, Meghreblian B, Williams JMG. Cyberchondriasis: fact or fiction? A preliminary examination of the relationship between health anxiety and searching for health information on the Internet. J Anxiety Disord 2012 Jan;26(1):189-196. [doi: 10.1016/j.janxdis.2011.11.005] [Medline: 22137465]
44. Ryan A, Wilson S. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin Drug Saf 2008 May;7(3):227-229. [doi: 10.1517/14740338.7.3.227] [Medline: 18462181]
45. Tan S, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
46. Hernández-García I, Giménez-Júlvez T. Assessment of health information about COVID-19 prevention on the internet: infodemiological study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]
47. Cuan-Baltazar J, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
48. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
49. Su C, Ng H, Yang A, Lin C. Psychometric evaluation of the Short Form 36 Health Survey (SF-36) and the World Health Organization Quality of Life Scale Brief Version (WHOQOL-BREF) for patients with schizophrenia. Psychol Assess 2014 Sep;26(3):980-989. [doi: 10.1037/a0036764] [Medline: 24796341]

Abbreviations

ANOVA: analysis of variance
CFI: comparative fit index
HBM: health belief model
PMT: protection motivation theory
PPE: Personal protective equipment
RMSEA: root mean square error of approximation
SEM: structural equation modeling

SRMR: standardized root mean square residual
STAI: State-Trait Anxiety Inventory
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 21.09.20; peer-reviewed by A Al-Hasan, MA Bahrami; comments to author 03.10.20; revised version received 19.10.20; accepted 28.10.20; published 24.12.20.

Please cite as:
Pramukti I, Strong C, Sitthimongkol Y, Setiawan A, Pandin MGR, Yen CF, Lin CY, Griffiths MD, Ko NY
Anxiety and Suicidal Thoughts During the COVID-19 Pandemic: Cross-Country Comparative Study Among Indonesian, Taiwanese, and Thai University Students
J Med Internet Res 2020;22(12):e24487
URL: http://www.jmir.org/2020/12/e24487/
doi:10.2196/24487
PMID:33296867

Original Paper

Use of Asynchronous Virtual Mental Health Resources for COVID-19 Pandemic–Related Stress Among the General Population in Canada: Cross-Sectional Survey Study

Chris G Richardson[1], PhD; Allie Slemon[2], RN, MSN; Anne Gadermann[1,3], PhD; Corey McAuliffe[1], PhD; Kimberly Thomson[1,3], PhD; Zachary Daly[2], RN, MS; Travis Salway[4], PhD; Leanne M Currie[2], RN, PhD; Anita David[5], BA; Emily Jenkins[2], RN, MPH, PhD

[1]Centre for Health Evaluation and Outcome Sciences, Providence Healthcare Research Institute, St Paul's Hospital, Vancouver, BC, Canada
[2]School of Nursing, University of British Columbia, Vancouver, BC, Canada
[3]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
[4]Simon Fraser University, Burnaby, BC, Canada
[5]Patient Voices Network, Vancouver, BC, Canada

Corresponding Author:
Emily Jenkins, RN, MPH, PhD
School of Nursing
University of British Columbia
T201-2211 Wesbrook Mall
Vancouver, BC, V6B 2B5
Canada
Phone: 1 604 822 4980
Email: emily.jenkins@ubc.ca

Abstract

Background: The COVID-19 pandemic has resulted in profound mental health impacts among the general population worldwide. As many in-person mental health support services have been suspended or transitioned online to facilitate physical distancing, there have been numerous calls for the rapid expansion of asynchronous virtual mental health (AVMH) resources. These AVMH resources have great potential to provide support for people coping with negative mental health impacts associated with the pandemic; however, literature examining use prior to COVID-19 illustrates that the uptake of these resources is consistently low.

Objective: The aim of this paper is to examine the use of AVMH resources in Canada during the COVID-19 pandemic among the general population and among a participant subgroup classified as experiencing an adverse mental health impact related to the pandemic.

Methods: Data from this study were drawn from the first wave of a large multiwave cross-sectional monitoring survey, distributed from May 14 to 29, 2020. Participants (N=3000) were adults living in Canada. Descriptive statistics were used to characterize the sample, and bivariate cross-tabulations were used to examine the relationships between the use of AVMH resources and self-reported indicators of mental health that included a range of emotional and coping-related responses to the pandemic. Univariate and fully adjusted multivariate logistic regression models were used to examine associations between sociodemographic and health-related characteristics and use of AVMH resources in the subgroup of participants who reported experiencing one or more adverse mental health impacts identified in the set of self-reported mental health indicators.

Results: Among the total sample, 2.0% (n=59) of participants reported accessing AVMH resources in the prior 2 weeks to cope with stress related to the COVID-19 pandemic, with the highest rates of use among individuals who reported self-harm (n=5, 10.4%) and those who reported coping "not well" with COVID-19–related stress (n=22, 5.5%). Within the subgroup of 1954 participants (65.1% of the total sample) who reported an adverse mental health impact related to COVID-19, 54 (2.8%) reported use of AVMH resources. Individuals were more likely to have used AVMH resources if they had reported receiving in-person mental health supports, were connecting virtually with a mental health worker or counselor, or belonged to a visible minority group.

Conclusions: Despite substantial government investment into AVMH resources, uptake is low among both the general population and individuals who may benefit from the use of these resources as a means of coping with the adverse mental health impacts of the COVID-19 pandemic. Further research is needed to improve our understanding of the barriers to use.

(J Med Internet Res 2020;22(12):e24868) doi:10.2196/24868

KEYWORDS

virtual health; digital health; virtual mental health; mental health; public health; COVID-19; coping; stress; implementation; utilization

Introduction

With growing global access to the internet [1], the World Health Organization [2] has identified digital health as a key health promotion strategy that can “benefit people in a way that is ethical, safe, secure, reliable, equitable and sustainable." Broadly, digital mental health comprises health care activities that are not carried out in traditional face-to-face format and may include synchronous (ie, in real time) video visits with mental health care providers or asynchronous (ie, not in real time) use of online self-management skill-building programs [3]. As a core aspect of the growth in digital health, there has been a recent surge in free, publicly available virtual mental health resources, including online programs and smartphone apps. In this paper, we use the term *asynchronous virtual mental health* (AVMH) *resources* to refer to any app, website, online tool, or other online support that does *not* involve synchronous direct contact with a mental health care provider. Although many of these publicly available mental health apps and online programs have been endorsed by users with generally high ratings and positive review comments [4], most are lacking direct scientific evidence of their effectiveness [5]. While not a replacement for professional decision making, AVMH resources nonetheless offer an alternative to in-person engagement with a mental health clinician, and aim to support people coping with symptoms related to a wide range of mental health conditions and challenges, including mild to moderate depression and anxiety, trauma, substance use, and other psychosocial stressors [6,7]. While limited, the research that has been conducted on smartphone and internet-based resources has demonstrated that AVMH resources can be effective in supporting people to manage their symptoms [8,9] and may reduce the need for clinical care [6].

The COVID-19 pandemic presents a heightened opportunity for AVMH resources to support mental well-being. The mental health impacts of COVID-19 are profound, including significant increases in depression, anxiety, and stress among the general population [10-13]. Public health measures, such as lockdown and quarantine, are necessary for reducing the spread of the virus but are also associated with negative psychological effects [14,15]. Loss of employment, anxiety about physical health and COVID-19 exposure, and social isolation are additional stressors that can contribute to poor mental health during the pandemic and beyond [16-18]. Further, researchers have warned of a likely increase in self-harm and suicide related to COVID-19 stressors [19,20]. In May 2020, many in-person mental health supports were reduced, suspended, or transitioned to online services in Canada, amid numerous reports from researchers around the globe calling for the rapid expansion of AVMH resources to address the mental health impact of the pandemic [21-24]. In the context of limited in-person services, AVMH resources represent a scalable means of promoting coping among those experiencing mental health challenges, including members of the general population who have not engaged with the mental health system but would likely benefit from these resources as a means of managing the mental health impacts of the COVID-19 pandemic [8,9,23].

Despite the potential for AVMH resources to support coping with COVID-19–related stressors, research suggests that uptake has been consistently low. While research examining AVMH resources in the context of COVID-19 is still emerging and no Canadian data are currently available, Yao and colleagues [25] reported results from a survey in China indicating that only 3.7% of respondents experiencing high levels of acute stress had used mental health services in any form in the early months of the COVID-19 outbreak. To our knowledge, few, if any, studies have examined the uptake of AVMH resources in the COVID-19 context. Prior to COVID-19, AVMH uptake and retention have been a perpetual challenge, with many internet-based resources and apps having very low use among the general population [26,27]. Moreover, use of AVMH resources tends to decline sharply in the first week following initial use; for example, Baumel and colleagues [28] found that as few as 4% of users of mental health apps continued to access the resource 15 days after initial installation.

In Canada, various AVMH programs have been promoted to support individuals’ mental health during the COVID-19 pandemic, including BounceBack [29], Wellness Together Canada [30], WellCan [31], and MindBeacon [32]. Free and available to the public, these resources hold promise for improving mental health and well-being at a time when in-person services are reduced and travel for services is limited, particularly for individuals in rural and remote areas. While a digital divide persists, particularly among low-income and rural populations [33], internet access is widespread in Canada, supporting the uptake of AVMH resources. In 2018, 94% of Canadians reported home internet access, and 88% of internet users had a smartphone [34]. AVMH resources have great potential to promote mental health and coping with stressors during the COVID-19 pandemic, given their demonstrated efficacy in supporting individuals’ self-management of a variety of mental health challenges [8,35]; however, little is known about the uptake of these resources among the general public in Canada in the pandemic context. Further, as mentioned, the global evidence on the use of similar virtual resources remains extremely limited. The aim of this investigation is to examine the use of AVMH resources during the COVID-19 pandemic among the general population in Canada to inform intervention research, development, and refinement of asynchronous virtual public health solutions related to mental health service delivery.

Methods

Overview

Data for this analysis were drawn from the first wave of a large multiwave cross-sectional monitoring survey, *Assessing the Impacts of COVID-19 on Mental Health*. This study utilizes a nationally representative sample of people aged ≥18 years living in Canada with the aim to examine mental health, stress, and coping during the COVID-19 pandemic. This project involves a unique partnership between academic researchers and 2 not-for-profit mental health organizations, the Canadian Mental Health Association (CMHA) and the Mental Health Foundation in the United Kingdom. Survey items were largely drawn from a repeated cross-sectional survey first commissioned by the Mental Health Foundation for COVID-19. Adaptations were made for the Canadian context.

Data Collection

Wave 1 of the survey was distributed from May 14 to 29, 2020, by Maru/Matchbox, a national polling vendor that maintains the online Maru Voice Panel of 125,000 Canadian adults. Members of the panel were randomly invited to participate in the survey from national census–informed stratifications defined by sociodemographic characteristics (age, gender, household income, and region), with adjustments for response propensity. Maru/Matchbox utilized target sampling to ensure inclusion of populations that may be difficult to reach via the internet, including older adults and racialized populations. In total, 3558 participants were invited by email to complete the survey to yield a total of 3000 respondents (84% invitation-to-response rate). The survey was available in English and French, Canada’s two official languages. The data collection time period was chosen to correspond with the beginning phases of the reopening of many of Canada’s provinces and territories, following approximately 2 months of public health orders that resulted in closures and physical distancing after COVID-19 was declared a pandemic by the World Health Organization on March 11, 2020 [36]. Additional details on the survey development and implementation have recently been published [13].

Measures and Analysis

Participants reported on their sociodemographic information as well as their mental health during the COVID-19 pandemic. Participants were asked to rate their mental health *now* compared to *before* the COVID-19 pandemic and related restrictions. Participants also reported on their overall coping and emotional responses to the COVID-19 pandemic in the past 2 weeks, including feeling depressed, stressed, hopeless, and anxious. Finally, participants reported if they had experienced suicidal thoughts/feelings or deliberate self-harm as a result of the COVID-19 pandemic in the past 2 weeks. These indicators were informed by research on previous pandemics with further refinement carried out via a citizen’s jury participatory methodology process involving people with lived experience of mental health conditions [37].

To assess the use of AVMH resources during COVID-19, participants responded to the survey item, “Which of the following have helped you cope with stress related to the COVID-19 pandemic in the past 2 weeks?” Participants were asked to select all that apply from a checklist of items that included three mental health strategies: (1) accessing virtual mental health resources (eg, online cognitive behavioral therapy, etc); (2) connecting with a mental health worker or counselor virtually (eg, via phone, video chat, etc); and (3) receiving in-person mental health supports. Study participants who endorsed the first option were classified as having used AVMH resources for this investigation.

Descriptive statistics were used to characterize the sample, and bivariate cross-tabulations were used to examine the relationships between self-reported indicators of mental health and the use of AVMH resources. Any participant who reported experiencing a negative impact of the pandemic on one or more of the self-reported mental health indictors was classified as experiencing an adverse mental health impact. The subgroup of participants who reported experiencing an adverse mental health impact included participants who reported that their mental health was now “slightly worse” or “significantly worse” compared to before the COVID-19 pandemic; participants who indicated that their level of coping with the stress of the COVID-19 pandemic was “not very well” or “not well at all”; and participants who reported feeling panicked, anxious or worried, hopeless, stressed, or depressed as a result of the COVID-19 pandemic. Additionally, respondents who reported experiencing suicidal thoughts/feelings as a result of the COVID-19 pandemic in the past 2 weeks and those who reported deliberately hurting themselves as a result of the COVID-19 pandemic in the past 2 weeks were also classified as experiencing an adverse mental health impact. Univariate and fully adjusted multivariate logistic regression models were then used to examine the associations between a set of a priori specified sociodemographic and health-related characteristics and use of AVMH resources in the subgroup of participants classified as experiencing an adverse mental health impact.

Ethics

Ethical approval for this study was provided by the Behavioural Research Ethics Board at the University of British Columbia (H20-01273). Participants provided consent online prior to beginning the survey. Maru/Matchbox provided all participants with a small honorarium for completing the survey.

Results

The sample comprised 3000 participants; detailed sociodemographic characteristics are presented in Table 1.

Table 1. Sociodemographic characteristics of the sample.

Characteristic	Total respondents (N=3000), n (%)[a]
Gender[b]	
Male	1467 (48.9)
Female	1533 (51.1)
Age group (years)	
Youth (18-24)	62 (2.1)
Adult (25-64)	2145 (71.5)
Senior (≥65)	793 (26.4)
Household income ($CAD)	
<$25,000	234 (7.8)
$25,000-$50,000	504 (16.8)
$50,000-$100,000	992 (33.1)
>$100,000	1270 (42.3)
Education	
High school or less	431 (14.4)
Some university or college	1139 (38.0)
University or college degree (diploma, undergraduate, and/or graduate degree)	1430 (47.7)
Employment (multiple responses permitted)	
Working full time (≥30 hours per week)	1225 (40.8)
Working part time (<30 hours per week)	286 (9.5)
Retired	882 (29.4)
Full-time student (eg, school, college, university, job training)	50 (1.7)
Part-time student (eg, school, college, university, job training)	16 (0.5)
Unemployed (due to COVID-19)	284 (9.5)
Unemployed (prior to COVID-19)	103 (3.4)
Ethnic minority[c]	
Not a visible minority	2128 (70.9)
Visible minority	389 (13.0)
Indigenous (eg, First Nations, Inuit, Métis)	87 (2.9)
Province	
British Columbia/territories	440 (14.7)
Alberta	333 (11.1)
Ontario	1140 (38.0)
Quebec	658 (21.9)
Manitoba/Saskatchewan	194 (6.5)
Atlantic provinces	235 (7.8)
Location	
Urban	2516 (83.9)
Rural	484 (16.1)
Self-reported prior mental health condition (yes)	546 (18.2)
Accessed virtual mental health resources in the past 2 weeks (yes)	59 (2.0)

[a] Note that some category percentages do not add up to 100 due to missing responses.

[b] Maru/Matchbox, the polling vendor that distributed this survey, provides demographic data for each panel member collected prior to survey completion. Though this binary variable was used in this analysis, the research team recognizes that binary gender identities do not accurately reflect many individuals' self-identified gender.

[c] Visible minority and non–visible minority categories were constructed by the research team based on participants' self-reported ethnicity (eg, South Asian, Middle Eastern, European origins).

Among the total sample, 59 (2.0%) participants reported accessing AVMH resources in the prior 2 weeks to cope with stress related to the COVID-19 pandemic. Table 2 presents rates of self-reported AVMH resource use or nonuse across various self-reported mental health indicators. The highest rate of access was among individuals who reported having deliberately hurt themselves (n=5, 10.4%), those who reported coping “not well” with stress related to the COVID-19 pandemic (n=22, 5.5%), those who experienced suicidal thoughts/feelings (n=14, 8.0%), and those who reported having “worse” mental health compared to before the COVID-19 pandemic (n=40, 3.6%). There was also significant overlap across these subgroups in that 74.3% (n=130) of participants who reported experiencing suicidal thoughts/feelings and 54.2% (n=26) of participants who reported deliberately hurting themselves also reported having worse mental health now compared to before the pandemic. Though higher than the sample average, accessing AVMH resources remained low among those who experienced negative emotions related to COVID-19, including individuals who felt hopeless (n=16, 4.5%), depressed (n=29, 4.3%), panicked (n=10, 4.3%), stressed (n=32, 2.9%), and anxious (n=38, 2.8%).

To identify sociodemographic and health-related characteristics associated with reported use of AVMH resources, a subgroup was created that contained any participant who reported experiencing one or more adverse mental health impacts on the mental health indicators listed in Table 2. This subgroup comprised 1954 participants, of which 54 (2.8%) reported use of AVMH resources. Table 3 presents the results of a multivariate logistic regression model examining the association between sociodemographic characteristics and the use of AVMH resources within this subsample of participants who reported adverse mental health outcomes due to the COVID-19 pandemic.

Results of the fully adjusted multivariate logistic regression model indicate that, among the subsample who reported adverse mental health impacts due to COVID-19, those who reported receiving in-person mental health supports or connecting virtually with a mental health worker or counselor were significantly more likely to report accessing AVMH resources (odds ratio [OR] 6.05, 95% CI 1.35-27.17 and OR 8.96, 95% CI 4.36-18.42, respectively). Additionally, individuals belonging to visible minority groups were more likely to report accessing AVMH resources (OR 3.79, 95% CI 1.83-7.86).

Table 2. Accessing asynchronous virtual mental health (AVMH) resources across indicators of self-reported mental health impact.

Mental health indicator	Used AVMH resource, n (%)	Did not use AVMH resource, n (%)
Compared to before the COVID-19 pandemic and related restrictions in Canada, how would you say your mental health is now?[a]		
About the same or better (n=1874)	19 (1.0)	1855 (99.0)
Worse (n=1121)	40 (3.6)	1081 (96.4)
Overall, how well do you think you are coping with stress related to the COVID-19 pandemic?[a]		
Fairly or very well (n=2439)	36 (1.5)	2403 (98.5)
Not well (n=400)	22 (5.5)	378 (94.5)
Felt panicked as a result of the COVID-19 pandemic in the past 2 weeks[b]		
Yes (n=231)	10 (4.3)	221 (95.7)
No (n=2769)	49 (1.8)	2720 (98.2)
Felt anxious or worried as a result of the COVID-19 pandemic in the past 2 weeks[a]		
Yes (n=1361)	38 (2.8)	1323 (97.2)
No (n=1639)	21 (1.3)	1618 (98.7)
Felt hopeless as a result of the COVID-19 pandemic in the past 2 weeks[a]		
Yes (n=353)	16 (4.5)	337 (95.5)
No (n=2647)	43 (1.6)	2604 (98.4)
Felt stressed as a result of the COVID-19 pandemic in the past 2 weeks[a]		
Yes (n=1103)	32 (2.9)	1071 (97.1)
No (n=1897)	27 (1.4)	1870 (98.6)
Felt depressed as a result of the COVID-19 pandemic in the past 2 weeks[a]		
Yes (n=676)	29 (4.3)	647 (95.7)
No (n=2324)	30 (1.3)	2294 (98.7)
Experienced suicidal thoughts/feelings in the past 2 weeks[a]		
Yes (n=176)	14 (8)	162 (92)
No (n=2792)	43 (1.5)	2749 (98.5)
Deliberately hurt myself in the past 2 weeks[a]		
Yes (n=48)	5 (10.4)	43 (89.6)
No (n=2936)	52 (1.8)	2884 (98.2)

[a] *P*<.01, based on chi-square analyses or exact test (Fischer or Fisher-Freeman-Halton).

[b] *P*<.05, based on chi-square analyses or exact test (Fischer or Fisher-Freeman-Halton).

Table 3. Results of univariate and fully adjusted multivariate logistic regression models examining the association between use of asynchronous virtual mental health resources and sociodemographic characteristics among participants who reported adverse mental health impacts related to the COVID-19 pandemic (n=1954).

Variable	Unadjusted odds ratio (95% CI)	Adjusted odds ratio (95% CI)
Age group (reference: youth)		
Adult	0.70 (0.16-2.96)	1.13 (0.24-5.44)
Senior	0.27 (0.05-1.41)	0.31 (0.04-2.67)
Gender (female)	1.23 (0.71-2.14)	1.38 (0.70-2.74)
LGBTQ2+[a] (yes)	2.09 (0.97-4.52)	1.42 (0.58-3.50)
Income (reference: >$100,000)		
<$25,000	1.49 (0.59-3.78)	0.63 (0.17-2.32)
$25,000-$50,000	1.33 (0.61-2.87)	0.84 (0.31-2.25)
$50,000-$100,000	1.16 (0.61-2.21)	1.01 (0.48-2.10)
Visible minority (reference: no)		
Yes	3.59 (1.95-6.61)	3.79 (1.83-7.86)[b]
Indigenous	1.79 (0.42-7.73)	1.69 (0.35-8.28)
Location (rural)	0.69 (0.29-1.63)	1.08 (0.39-3.04)
Highest level of education completed (reference: university)		
High school or less	0.41 (0.12-1.36)	0.52 (0.14-1.98)
Some university or college	1.07 (0.61-1.87)	1.01 (0.50-2.03)
Disability (yes)	1.75 (0.86-3.52)	1.38 (0.54-3.53)
Pre-existing mental health condition (yes)	2.97 (1.72-5.12)[b]	1.69 (0.80-3.58)
Receiving in-person mental health supports (yes)	11.61 (3.66-36.87)[b]	6.05 (1.35-27.17)[c]
Connecting with a mental health worker or counselor virtually (eg, via phone, video chat, etc) (yes)	12.19 (6.68-22.24)[b]	8.96 (4.36-18.42)[b]

[a] LGBTQ2+: lesbian, gay, bisexual, transgender, queer/questioning, and two-spirited.

[b] *P*<.01.

[c] *P*<.05.

Discussion

Principal Findings

Findings from this study illustrate that the reported uptake of AVMH resources was extremely low (2.0%) among the general population in Canada in the context of the first 3 months of the COVID-19 pandemic. A similarly low level of use (2.8%) was observed in the subgroup of participants who reported an adverse mental health impact related to COVID-19. While utilization of these resources was higher among particular groups experiencing mental health challenges as a result of the pandemic, including those who had engaged in self-harm behaviors, experienced suicidal thoughts, or reported poor coping, uptake remained low even among these groups. While representing a different national context for mental health care, these findings are consistent with Yao and colleagues' [25] report of very low use of any mental health services among adults in China experiencing high levels of acute stress during COVID-19. Despite evidence indicating efficacy of select resources [38,39], promotion of existing AVMH resources (eg, [40]), and numerous calls for the expansion of these resources among researchers and public health officials, there is a dearth of evidence regarding the uptake of AVMH resources among the general public and population subgroups experiencing adverse mental health impacts during the current pandemic. This study is among the first to present data regarding the extent to which members of the general public report accessing AVMH resources during the COVID-19 pandemic, while also examining access among a subgroup of individuals experiencing adverse mental health impacts related to the pandemic. In this section, we discuss key findings from this study: the low uptake of AVMH resources within the sample, higher use among the visible ethnic minority subgroup and those who accessed other forms of mental health support, and limited use of AVMH resources among those who reported self-harm or suicidal thoughts.

COVID-19 has presented unexpected and significant stressors that have been demonstrated to negatively impact mental health worldwide [10,12]. While for some individuals, the mental health impacts are considerable, including posttraumatic stress disorder, self-harm, and suicidal ideation [20,41], most people will experience mild to moderate symptoms of stress, anxiety, and low mood [11,42,43]. Such subclinical symptoms have been shown to be particularly responsive to self-directed mental health interventions, including virtual resources [44]. Research also indicates that accessing internet-based cognitive behavioral therapy for depression is associated with reductions in suicidal ideation [45], suggesting that use of AVMH resources can also contribute to reductions in these adverse mental health experiences. In the context of COVID-19 public health measures, including lockdown and physical distancing, mental health apps and online programs are particularly important public resources that are free or low cost, widely available, and can support mental health without additional risks for spreading the virus. However, research demonstrates that prior to COVID-19, uptake of AVMH resources was consistently low [26,28]. Our findings illustrate that despite the increased value of virtual supports during the pandemic and promotion of AVMH resources in Canada, use of these resources remains extremely limited within the general population. Concerningly, use remains low (<10.5%) among population subgroups who would appear to benefit from their use, including those experiencing suicidal thoughts, anxiety, hopelessness, stress, and depression related to COVID-19. While many companies and organizations are producing AVMH resources, there is an expectation of “build it and they will come,” but that does not appear to be the case.

Previous literature on the challenges of AVMH resources sheds light on potential reasons for poor uptake in the context of COVID-19. Many users have reported technological issues that hinder the usability of the apps, including poor speed or stability of their internet connection [46]. Additionally, potential users may be deterred by privacy and confidentiality concerns [6,47] or by a lack of comfort and familiarity with the internet and smartphone devices, such as among older adults [48]. While the AVMH resources promoted by governments and public health officials in Canada to support mental health during COVID-19 are evidence based and reputable, many apps advertised online or through app stores are not grounded in evidence or validated through research and may not actually be effective for managing mental health challenges [6,22]. Further research is needed to examine the barriers to widespread uptake of AVMH resources, particularly among individuals experiencing mental health impacts due to the COVID-19 pandemic. Such research can guide refinement of existing AVMH resources and promotion strategies for these resources among the general public and those experiencing mental health challenges.

In addition to presenting findings on accessing of virtual mental health supports among the general population, findings from this study illustrate that a higher proportion of specific population subgroups reported using these resources. We found that belonging to a visible minority ethnic group was associated with an increased likelihood of accessing AVMH resources, even after adjusting for age, income, and education. A potential factor in increased use among these populations may be heightened stigma toward mental health challenges within certain cultures. For example, cultural values held by some individuals of Asian descent have contributed to higher stigma about mental health challenges and reduced help seeking for in-person mental health supports [49,50]. Research has further suggested that while individuals of Asian descent may exhibit fewer mental health help-seeking behaviors overall, there is evidence of an increased preference among Asian American college students (compared to Caucasian students) for online-only supports versus face-to-face services [51]. It is possible that the anonymity of virtual mental health supports may therefore facilitate increased use among particular visible minority populations, as a preferred alternative to seeking one-on-one or group support from a mental health clinician.

Additionally, individuals who were either receiving in-person mental health supports or connecting virtually with a mental health worker or counselor were significantly more likely to report having accessed AVMH supports. This suggests that those who seek support may be motivated to try multiple strategies for supporting their own mental health, or that mental health clinicians may be directing clients toward AVMH resources to augment individual therapy. Previous literature demonstrates that integrating the use of virtual resources with other forms of mental health care, such as discussing trends from a mental health app with a clinician during an in-person appointment, can lead to improved outcomes [52]. As such, enhancing clinicians’ awareness of AVMH resources and providing training specifically for clinicians on how to integrate these resources into mental health care, may increase uptake among individuals who may benefit from the support they provide. Research has further demonstrated that while most users do not sustain engagement after initial sign-up [27], supports built into the system to facilitate engagement can be effective. Specifically, uptake and retention are significantly higher among users of AVMH resources who receive peer support [27], and automated conversational agents such as chatbots have been shown to effectively support sustained engagement [53]. While clinician and peer support may be logistically challenging and costly, Burger and colleagues [54] note that current AVMH resources “do not get close to the full technological potential of e-mental health” and that enhancing the available features of these resources may lead to greater uptake and retention of users.

While AVMH resources were underutilized across the entire sample, including among those experiencing mental health challenges, what was particularly concerning was the relatively low uptake among those who had deliberately self-harmed or had suicidal thoughts/feelings in the past 2 weeks. While specific AVMH resources related to suicide are fewer compared to those addressing mental health challenges such as low mood and anxiety [55], many resources address a broad range of mental health concerns that may underlie self-harm or suicidal ideation. Although a recent systematic review and meta-analysis concluded that internet-based interventions directly targeting suicidal ideation using cognitive behavioral therapy represent a promising means of providing low-threshold support to address suicidal ideation [56], researchers have suggested that AVMH resources may not represent the most appropriate form of mental health support for individuals with this experience given the severity of symptoms and degree of risk involved [57]. Despite relatively low uptake, use of mental health resources among individuals who reported self-harm or suicidal ideation was significantly higher compared to the general population, suggesting a desire for support seeking in this population. As Frost and Casey [58] suggest, utilizing virtual resources may be a critical stepping stone toward engaging in more intensive in-person clinical support.

Limitations

The sampling approach used in this study aimed to generate a large nationally representative sample of individuals living in Canada. The representativeness of the sample was supported since, according to the results from the 2016 Canadian census [59], 77.7% of the population did not belong to a visible minority (versus 70.9% in our sample), and 18.7% of the population was considered rural (versus 16.1% in our sample). In terms of household income, 57.7% of individuals reported an annual household income that was less than $100,000, which is consistent with recent data from Statistics Canada indicating that the average household income in Canada in 2017 was $93,300 [60]. However, while the sample was representative in these ways, there are other ways in which it did not reflect the overall population of Canada. For example, because it was an online survey, only individuals with internet access were able to participate. Individuals who were unable to afford the cost of internet service or equipment, or those living in rural "blackout zones" are therefore not represented in our sample [34,61]. As such, the actual use of AVMH resources among the general population may be even lower than that reflected in our findings.

The indicators of mental health used in the survey did not include clinical assessments of mental illness nor did it include prepandemic measures of baseline mental health. However, as noted in the description of study methods, the selection of items was informed by research on previous pandemics and the items were refined in consultation with people with lived experience of mental health conditions through a citizen's jury participatory methodology process [37]. Additionally, accessing AVMH resources was assessed in this survey through a single item measure. We are therefore unable to present data on which AVMH resources were used (eg, websites, apps, podcasts), participants' perceptions of the effectiveness of the resources, length of use, or why they may have stopped using these resources. Additionally, our survey asked participants about accessing AVMH resources to cope with COVID-19 stress and may not have captured use of these resources for other reasons, such as to help manage other mental health challenges. To further examine the use of AVMH resources among people living in Canada, we have expanded our survey content for subsequent waves of data collection to add further depth and nuance to the results presented in this investigation. It is also important to note that the cross-sectional design of this study prevents us from making causal conclusions. However, we will be conducting future waves of data collection over the course of the pandemic that will help address this limitation.

Comparison With Prior Work

While multiple researchers have highlighted the importance of virtual resources to address the significant mental health impacts of the COVID-19 pandemic, there remains a paucity of data on their use, uptake, and retention. However, previous research suggests that these resources may be underutilized [28], and emerging data within the COVID-19 context suggest low uptake among acutely stressed members of the general population [25]. Our examination of the use of AVMH resources related to COVID-19 stressors among people living in Canada is one of the first to provide empirical evidence on trends in accessing AVMH resources among the general population, with particular attention to their uptake among people experiencing adverse mental health impacts related to the current COVID-19 pandemic. Our analysis illustrates that not only is overall use of AVMH resources low, but use remains low even among individuals who would be predicted to benefit from these resources, such as individuals experiencing negative emotions (eg, anxiety) related to COVID-19 and individuals experiencing worsening mental health and difficulty coping.

Conclusions

AVMH resources hold great promise to support mental well-being during the COVID-19 pandemic by helping individuals cope with stressors and negative mental health impacts. However, despite efforts to promote AVMH resources within Canada, the results from this study suggest that many individuals who may benefit from these resources are not accessing them to help manage mental health challenges. Further research that examines the barriers and facilitators to AVMH resource use is needed to enhance uptake and impact.

Acknowledgments

The authors thank the CMHA and the Mental Health Foundation for supporting this monitoring study. We are grateful to the CMHA for funding Maru/Matchbox to deploy the survey. Special thanks to Jonathan Morris, Katherine Janson, Andrew Galley, and Margaret Eaton (CMHA) for assisting with study communications and government relations outreach, and to Neesha Mathew and Jacqueline Campbell (Maru/Matchbox) for managing survey deployment and data preparation.

EJ and AG both hold Scholar Awards from the Michael Smith Foundation for Health Research, which has supported their time contributions to the study. JM, employed by the CMHA, contributed to the development of survey content. The CMHA had no further role in the study design, data collection, data analysis, or interpretation.

Conflicts of Interest

CR reports receiving personal fees from the University of British Columbia during the conduct of this study. All other authors report no competing interests.

References

1. Individuals using the internet (% of population). The World Bank. 2020. URL: https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2018&start=1990 [accessed 2020-12-15]
2. Draft global strategy on digital health 2020-2025. World Health Organization. p. 5 URL: https://www.who.int/health-topics/digital-health [accessed 2020-12-15]
3. McGrail KM, Ahuja MA, Leaver CA. Virtual visits and patient-centered care: results of a patient survey and observational study. J Med Internet Res 2017 May 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
4. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep 10;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]
5. Neary M, Schueller SM. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537 [FREE Full text] [doi: 10.1016/j.cbpra.2018.01.002] [Medline: 33100810]
6. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone apps for the treatment of mental disorders: Systematic review. JMIR Mhealth Uhealth 2020 Apr 02;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
7. Radovic A, Vona PL, Santostefano AM, Ciaravino S, Miller E, Stein BD. Smartphone applications for mental health. Cyberpsychol Behav Soc Netw 2016 Jul;19(7):465-470 [FREE Full text] [doi: 10.1089/cyber.2015.0619] [Medline: 27428034]
8. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: A narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019 Jul 16;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
9. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
10. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020 Jul 06;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
11. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. Brain Behav Immun 2020 Oct;89:531-542 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
12. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
13. Jenkins E, McAuliffe C, Hirani S, Richardson C, Thomson K, McGuinness L, et al. A portrait of the early and differential mental health impacts of the COVID-19 pandemic (forthcoming). Preventive Medicine 2020.
14. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet 2020 Mar 14;395(10227):912-920. [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
15. Luo X, Estill J, Wang Q, Lv M, Liu Y, Liu E, et al. The psychological impact of quarantine on coronavirus disease 2019 (COVID-19). Psychiatry Res 2020 Sep 02;291(9):113193. [doi: 10.1016/j.psychres.2020.113193] [Medline: 32544708]
16. Newby JM, O'Moore K, Tang S, Christensen H, Faasse K. Acute mental health responses during the COVID-19 pandemic in Australia. PLoS One 2020 Jul 28;15(7):e0236562 [FREE Full text] [doi: 10.1371/journal.pone.0236562] [Medline: 32722711]
17. Palgi Y, Shrira A, Ring L, Bodner E, Avidor S, Bergman Y, et al. The loneliness pandemic: Loneliness and other concomitants of depression, anxiety and their comorbidity during the COVID-19 outbreak. J Affect Disord 2020 Oct 01;275:109-111 [FREE Full text] [doi: 10.1016/j.jad.2020.06.036] [Medline: 32658811]
18. Wilson J, Lee J, Fitzgerald H, Oosterhoff B, Sevi B, Shook NJ. Job insecurity and financial concern during the COVID-19 pandemic are associated with worse mental health. J Occup Environ Med 2020 Sep;62(9):686-691. [doi: 10.1097/JOM.0000000000001962] [Medline: 32890205]
19. Kawohl W, Nordt C. COVID-19, unemployment, and suicide. Lancet Psychiatry 2020 May;7(5):389-390 [FREE Full text] [doi: 10.1016/S2215-0366(20)30141-3] [Medline: 32353269]
20. McIntyre RS, Lee Y. Projected increases in suicide in Canada as a consequence of COVID-19. Psychiatry Res 2020 Aug;290:113104 [FREE Full text] [doi: 10.1016/j.psychres.2020.113104] [Medline: 32460184]
21. Gaebel W, Stricker J. E‐mental health options in the COVID‐19 pandemic and beyond. Psychiatry Clin. Neurosci 2020 Jun 28;74(8):441-442. [doi: 10.1111/pcn.13079]
22. Marshall JM, Dunstan DA, Bartik W. The role of digital mental health resources to treat trauma symptoms in Australia during COVID-19. Psychol Trauma 2020 Aug;12(S1):S269-S271. [doi: 10.1037/tra0000627] [Medline: 32496103]
23. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: The 'black swan' for mental health care and a turning point for e-health. Internet Interv 2020 Apr;20:100317 [FREE Full text] [doi: 10.1016/j.invent.2020.100317] [Medline: 32289019]

24. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The role of telehealth in reducing the mental health burden from COVID-19. Telemed J E Health 2020 Apr 23;26(4):377-379 [FREE Full text] [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
25. Yao H, Chen J, Xu Y. Rethinking online mental health services in China during the COVID-19 epidemic. Asian J Psychiatr 2020 Apr;50:102015 [FREE Full text] [doi: 10.1016/j.ajp.2020.102015] [Medline: 32247261]
26. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: Systematic review of real-world uptake and engagement with digital self-help interventions for depression, low Mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
27. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
28. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
29. BounceBack. Canadian Mental Health Association British Columbia. 2019. URL: https://bouncebackbc.ca/ [accessed 2020-12-15]
30. Wellness Together Canada. URL: https://ca.portal.gs/ [accessed 2020-12-15]
31. Morneau S. WellCan. 2020. URL: https://wellcan.ca/ [accessed 2020-12-15]
32. MindBeacon. Health tech innovator MindBeacon works with province of Ontario to help people overcome mental health impact of the COVID-19 crisis. CNW Group Ltd. 2020 May 5. URL: https://www.newswire.ca/news-releases/health-tech-innovator-mindbeacon-works-with-province-of-ontario-to-help-people-overcome-mental-health-impact-of-the-covid-19-crisis-885266748.html [accessed 2020-12-15]
33. Stewart B. How COVID-19 worsens Canada's digital divide. CBC News. 2020 Sep 23. URL: https://www.cbc.ca/news/canada/british-columbia/covid-19-highlights-urban-rural-digital-divide-1.5734167 [accessed 2020-12-15]
34. Canadian internet use survey. Statistics Canada. 2019. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/191029/dq191029a-eng.htm [accessed 2020-12-15]
35. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: Systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
36. Vogel L. COVID-19: a timeline of Canada's first-wave response. CMAJ News. 2020 Jun 12. URL: https://cmajnews.com/2020/06/12/coronavirus-1095847 [accessed 2020-12-15]
37. Kousoulis A, McDaid S, Crepaz-Keay D. The COVID-19 pandemic, financial inequality and mental health. Mental Health Foundation. URL: https://www.mentalhealth.org.uk/our-work/research/coronavirus-mental-health-pandemic/covid-19-inequality-briefing [accessed 2020-12-15]
38. Annual report 2018-2019. Canadian Mental Health Association British Columbia. 2019. URL: https://cmha.bc.ca/wp-content/uploads/2019/10/CMHABC_AnnualReport_2018-19_WEB.pdf [accessed 2020-12-15]
39. Farvolden P. BEACON beta trial confirms effectiveness. MindBeacon Software Inc. 2020 Oct 28. URL: https://www.mindbeacon.com/insights-for-leaders/beacon-beta-trial-confirms-effectiveness [accessed 2020-12-15]
40. Government of Canada connects Canadians with mental wellness supports during COVID-19. Health Canada. 2020 Apr 15. URL: https://www.canada.ca/en/health-canada/news/2020/04/government-of-canada-connects-canadians-with-mental-wellness-supports-during-covid-190.html [accessed 2020-12-15]
41. Forte G, Favieri F, Tambelli R, Casagrande M. The enemy which sealed the world: Effects of COVID-19 diffusion on the psychological state of the italian population. J Clin Med 2020 Jun 10;9(6):1802 [FREE Full text] [doi: 10.3390/jcm9061802] [Medline: 32531884]
42. da Silva ML, Rocha RSB, Buheji M, Jahrami H, Cunha KDC. A systematic review of the prevalence of anxiety symptoms during coronavirus epidemics. J Health Psychol 2020 Aug 23. [doi: 10.1177/1359105320951620] [Medline: 32830577]
43. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Aug;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
44. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018 Mar;4(3) [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]
45. Mewton L, Andrews G. Cognitive behaviour therapy via the internet for depression: a useful strategy to reduce suicidal ideation. J Affect Disord 2015 Jan 01;170:78-84. [doi: 10.1016/j.jad.2014.08.038] [Medline: 25233243]
46. Feijt M, de Kort Y, Bongers I, Bierbooms J, Westerink J, IJsselsteijn W. Mental health care goes online: Practitioners' experiences of providing mental health care during the COVID-19 pandemic. Cyberpsychol Behav Soc Netw 2020 Dec 18;23(12):860-864. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]
47. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
48. Gould CE, Hantke NC. Promoting technology and virtual visits to improve older adult mental health in the face of COVID-19. Am J Geriatr Psychiatry 2020 Aug;28(8):889-890 [FREE Full text] [doi: 10.1016/j.jagp.2020.05.011] [Medline: 32425468]

49. Han M, Pong H. Mental health help-seeking behaviors among asian american community college students: The effect of stigma, cultural barriers, and acculturation. Journal of College Student Development 2015;56(1):1-14. [doi: 10.1353/csd.2015.0001]
50. Kam B, Mendoza H, Masuda A. Mental health help-seeking experience and attitudes in Latina/o American, Asian American, Black American, and White American college students. Int J Adv Counselling 2018 Oct 12;41(4):492-508. [doi: 10.1007/s10447-018-9365-8]
51. Lungu A, Sun M. Time for a change: College students' preference for technology-mediated versus face-to-face help for emotional distress. Telemed J E Health 2016 Dec;22(12):991-1000. [doi: 10.1089/tmj.2015.0214] [Medline: 27223266]
52. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
53. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: Mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
54. Burger F, Neerincx MA, Brinkman W. Technological state of the art of electronic mental health interventions for major depressive disorder: Systematic literature review. J Med Internet Res 2020 Jan 20;22(1):e12599 p.18 [FREE Full text] [doi: 10.2196/12599] [Medline: 31958063]
55. de la Torre I, Castillo G, Arambarri J, López-Coronado M, Franco MA. Mobile apps for suicide prevention: Review of virtual stores and literature. JMIR Mhealth Uhealth 2017 Oct 10;5(10):e130 [FREE Full text] [doi: 10.2196/mhealth.8036] [Medline: 29017992]
56. Büscher R, Torok M, Terhorst Y, Sander L. Internet-based cognitive behavioral therapy to reduce suicidal ideation: A systematic review and meta-analysis. JAMA Netw Open 2020 Apr 01;3(4):e203933 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3933] [Medline: 32343354]
57. Castillo-Sánchez G, Camargo-Henríquez I, Muñoz-Sánchez JL, Franco-Martín M, de la Torre-Díez I. Suicide prevention mobile apps: Descriptive analysis of apps from the most popular virtual stores. JMIR Mhealth Uhealth 2019 Aug 13;7(8):e13885 [FREE Full text] [doi: 10.2196/13885] [Medline: 31411144]
58. Frost M, Casey L. Who seeks help online for self-injury? Arch Suicide Res 2016 Feb 23;20(1):69-79. [doi: 10.1080/13811118.2015.1004470] [Medline: 25706352]
59. Data products, 2016 census. Statistics Canada. URL: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/index-eng.cfm [accessed 2020-12-15]
60. Real average total household income (before taxes), 2006-2017. Canada Mortgage and Housing Corporation. 2019 May 16. URL: https://www.cmhc-schl.gc.ca/en/data-and-research/data-tables/real-average-total-household-income-before-taxes [accessed 2020-12-15]
61. Emmanuel R. Canadians without reliable internet access being left behind, experts warn. iPolitics. 2020 May 8. URL: https://ipolitics.ca/2020/05/08/canadians-without-reliable-internet-access-being-left-behind-experts-warn/ [accessed 2020-12-15]

Abbreviations

AVMH: asynchronous virtual mental health
CMHA: Canadian Mental Health Association
OR: odds ratio

Edited by G Eysenbach; submitted 08.10.20; peer-reviewed by W Van Woensel, S Casteleyn, A Benis; comments to author 30.10.20; revised version received 27.11.20; accepted 11.12.20; published 30.12.20.

Please cite as:
Richardson CG, Slemon A, Gadermann A, McAuliffe C, Thomson K, Daly Z, Salway T, Currie LM, David A, Jenkins E
Use of Asynchronous Virtual Mental Health Resources for COVID-19 Pandemic–Related Stress Among the General Population in Canada: Cross-Sectional Survey Study
J Med Internet Res 2020;22(12):e24868
URL: http://www.jmir.org/2020/12/e24868/
doi:10.2196/24868
PMID:33315583

Viewpoint

Harnessing Digital Health Technologies During and After the COVID-19 Pandemic: Context Matters

Francesco Petracca[1*], MSc; Oriana Ciani[1,2*], MSc, PhD; Maria Cucciniello[3*], MSc, PhD; Rosanna Tarricone[1,4*], MSc, PhD

[1]Centre for Research in Health and Social Care Management (CERGAS), Government, Health and Non Profit Division, SDA Bocconi, Milan, Italy
[2]Institute of Health Research, University of Exeter Medical School, Exeter, United Kingdom
[3]University of Edinburgh Business School, Edinburgh, United Kingdom
[4]Department of Social and Political Sciences, Bocconi University, Milan, Italy
[*]all authors contributed equally

Corresponding Author:
Francesco Petracca, MSc
Centre for Research in Health and Social Care Management (CERGAS)
Government, Health and Non Profit Division
SDA Bocconi
Via Roentgen 1, 20136
Milan
Italy
Phone: 39 02 58365257
Email: francesco.petracca@unibocconi.it

Abstract

A common development observed during the COVID-19 pandemic is the renewed reliance on digital health technologies. Prior to the pandemic, the uptake of digital health technologies to directly strengthen public health systems had been unsatisfactory; however, a relentless acceleration took place within health care systems during the COVID-19 pandemic. Therefore, digital health technologies could not be prescinded from the organizational and institutional merits of the systems in which they were introduced. The Italian National Health Service is strongly decentralized, with the national government exercising general stewardship and regions responsible for the delivery of health care services. Together with the substantial lack of digital efforts previously, these institutional characteristics resulted in delays in the uptake of appropriate solutions, territorial differences, and issues in engaging the appropriate health care professionals during the pandemic. An in-depth analysis of the organizational context is instrumental in fully interpreting the contribution of digital health during the pandemic and providing the foundation for the digital reconstruction of what is to come after.

(J Med Internet Res 2020;22(12):e21815) doi:10.2196/21815

KEYWORDS

mobile apps; coronavirus; COVID-19; digital health; mHealth; organizational context; public health; telemedicine

The Unparalleled Surge in Digital Health Adoption

The COVID-19 pandemic has presented governments, managers, and professionals worldwide with unprecedented challenges, highlighting the limitations of analog health care systems that have long reinforced face-to-face models of care and congregation of individuals [1]. Regardless of the diverse containment and mitigation strategies implemented across countries, a common development observed during the pandemic is the renewed reliance on digital health, with efficient strategies implemented at various levels and directed at different stakeholders [2,3].

Rapid implementation of social distancing measures and rescheduling of elective procedures have led health care providers to resort to digital health applications for granting access to virtual consultations and remote visits and monitoring [4]. In short, health care has come to patients rather than the other way round.

As a result, during the first 3 months of the pandemic, an unparalleled surge in digital health adoption was observed, with a general scale-up of telemedicine [5] and a substantial shift to telehealth visits in some clinics [6], as confirmed by the increased number of telehealth services and e-consultations reported in several jurisdictions [6-8].

The fact that these technologies were available and ready for use should not have come as a surprise. However, before the COVID-19 outbreak, the uptake of digital health technologies to strengthen public health systems had been unsatisfactory, and the scholarly debate had rather focused on the barriers to adoption of digital health solutions [9,10] and difficulties in governing mobile apps [11]. Nonetheless, the adoption of digital health solutions, regardless of whether it occurred during or after the pandemic, cannot be prescinded from the organizational and institutional merits of the systems in which they are introduced. The full potential of technological innovation can be realized only by way of heightened attention to context.

COVID-19 Digital Health Experience in the Italian National Health Service

Italy was among the first and most significantly hit countries by SARS-CoV-2. As of July 21, 2020, Italy reported a total of 244,708 confirmed cases, over 34,000 COVID-19–related deaths, and the highest deaths per million inhabitants and case fatality rates worldwide [12-14].

In terms of institutional composition, the Italian National Health Service (SSN) is strongly decentralized and based on universal access to care. The central government exercises general stewardship, whereas 21 regions—each considerably different in size and economic development and autonomous in their health care management decisions—are responsible for the organization and delivery of primary, secondary, tertiary, and preventive health care services through local health authorities (LHAs) [15].

Before the COVID-19 emergency, adoption of digital health was a challenge in Italy, as elsewhere. By analyzing the distinctive digital health experiences during the first wave of the pandemic in Italy—a country that was on the front line of the pandemic [16,17], we aimed to showcase the best practices, open questions, and barriers encountered. Systems, institutions, and organizations who have attempted to embrace digital health in the past months and are looking to maintain the momentum after COVID-19 have experienced similar challenges. Some of these challenges are related to issues such as reimbursement of digital health services, experience of using contact tracing apps, the search for balance between local experiences and national stewardship, arduous involvement of all health care professionals, and the coexistence of digital and analog pathways.

Reimbursement Schemes

Appropriate reimbursement schemes are essential to expanding the role of digital technologies. During the current public health crisis, telemedicine services promptly seemed to be vital to support remote monitoring of less severe cases and ensure continuity of care for vulnerable, typically chronic patients, whose needs were overshadowed by the surge in patients with COVID-19 seeking assistance. However, telemedicine services in Italy have traditionally been scattered across a number of different applications, with poor interconnection and inconsistent local and regional reimbursement practices, since they are not specifically covered by the guaranteed health basket of the Italian SSN [18]. No extensive, nationwide input was provided by health authorities during the pandemic, as has been the case of France, where the Ministry of Health allowed the reimbursement of teleconsultations for suspected and confirmed COVID-19 cases [6]. The Italian National Institute of Health (Istituto Superiore di Sanità) proposed a temporary model to ease the implementation of telemedicine services during the emergency, but this essentially focused on the organizational requirements for telemedicine services and did not tackle the reimbursement domain [19]. The delay in the adoption of national guidance and specific reimbursement codes for telemedicine services resulted in inter-regional differences: jurisdictions such as Veneto, the Autonomous Province of Trento, and Tuscany promptly introduced digital health services as part of their discretionary guaranteed health benefits packages, and were followed by other pioneering regions in the following months. These packages are now being managed online with dedicated tariffs for the citizens of the above regions. Conversely, the remaining regions could only fund telemedicine services by implicitly allowing payment parity with standard outpatient procedures through regional discretionary spending. While the lack of a framework to pay for telehealth services hindered wider-scale adoption by public institutions, private telemonitoring service providers reported a marked increase in the use of direct-to-consumer services [18].

Contact Tracing Apps

Use of contact tracing techniques and apps during the pandemic exposed the challenges of safeguarding privacy in the development of prompt digital responses to address current and future needs. National contact tracing apps were identified as a key tool to tackle the epidemic and facilitate reopening of the economy. However, unsurprisingly, the design and adoption of these applications, along with associated evaluation timelines, have raised significant privacy and security concerns and resulted in considerable delays in their deployment during the pandemic. Like most European countries, Italy initially backed the adoption of centralized management and maintenance of contact tracing data, including the "social graph" of individuals a person has physically met over a given period [20]. This practice was in contrast with the view of the European Commission, which supported anonymous, aggregated, and decentralized approaches to contact tracing to preserve privacy [21]. Meanwhile, Apple and Google had collaborated in a novel partnership to develop a safe infrastructure and privacy protection standards to specifically tackle these operations [22]. However, voluntary, data minimization systems are considered the sole option to build acceptance and maximize the uptake of these solutions [23]. These systems preserve user anonymity while providing useful warnings to other users who may have recently come into close physical proximity of a certain individual who subsequently tested positive for COVID-19. Eventually, together with several other European countries [24], the Italian government shifted focus towards these types of

Bluetooth Low Energy apps that interfere as little as possible with users' privacy rights.

Collectively, lengthy negotiations on privacy standards and relatively scarce previous attempts to evaluate these technologies have delayed the widespread roll-out of functional contact-tracing apps.

Immuni, an app selected by an adhoc task force of the Italian Ministry of Health [25], was launched nationwide on June 15, 2020, that is, 6 weeks after the first lockdown in Italy was lifted (May 4, 2020). Despite significant setbacks that delayed the app deployment, Italy was still among the first European countries to roll-out a contact-tracing solution, as proof of the broad-based intricacy of this matter. Nonetheless, the number of app downloads (4.3 million) for Immuni has been disappointingly low so far and is insufficient to guarantee the effectiveness of the tool [26].

Balancing Local Experiences and National Responsibilities

Leveraging local experiences with digital health projects in response to the crisis must be balanced with the national governments' global policy-setting responsibilities. Highly regionalized health systems, such as the Italian SSN, can generate both significant opportunities for innovation as well as threats to national public health initiatives. Moreover, territorial differences in digital response to the COVID-19 crisis that have emerged could provide a basis for mutual learning. For instance, some regional health care systems were better prepared than others, not just in terms of telemedicine policies. In addition, in territories where large-scale digital health projects were ongoing, LHAs were able to rapidly apply them to the needs of the COVID-19 pandemic. For example, in 2015, the LHA of the Autonomous Province of Trento adopted TreC, a patient monitoring system that includes a mobile diary app and a web-based dashboard to facilitate patient-physician interaction and reduce direct access to hospitals. This system was designed to be interoperable with various electronic health records and was tested in several chronic populations [27-29]. After the COVID-19 outbreak in February 2020, the LHA, together with the research entity Fondazione Bruno Kessler, launched TreCovid19—an app built on the same infrastructure designed to inform citizens and enforce remote symptom monitoring. By the end of the pandemic peak, more than 9400 infected or quarantined cases from the province of Trento were actively monitored through this platform, with over 40,000 single parameters directly reported via the chatbot and approximately 700,000 individuals' information accessed via personal interviews with a local executive (data not reported elsewhere).

In the Puglia region in Southern Italy, similar work was being carried out for years on remote follow-up management of chronic patients through several pilot projects. This enabled the implementation of a remote monitoring system based on a web-based application for citizens, a web portal for health care professionals, and a telemedicine kit. The platform, called #Accasa, was implemented on April 22, 2020, for individuals under quarantine or those who have tested positive for COVID-19 [30]. Conversely, boundaries are sometimes blurred in decentralized systems when close cooperation among different institutional levels is required. Several related apps have proliferated in the market during the COVID-19 pandemic, often without the necessary interoperability described above and with overlapping or even hindering features concerning their proposed nationwide application. Although the central government was working on selecting and deploying the most appropriate solutions, all Italian regional systems have worked on their own apps. Initially, these apps only provided information to local communities; however, over time, extra functionalities were added, some of which were slated for inclusion in the design of the national contact-tracing app. For example, Friuli–Venezia Giulia in Northeast Italy had successfully completed the testing of the same app used for contact tracing in Singapore [31], whereas the island of Sardinia had actively sought solutions to track tourists in view of the upcoming summer holiday season. Both attempts were set aside, as contact-tracing needed to be managed nationally. Given that there was no guarantee that the regional apps would be interoperable with the national app, it was fundamental, but not trivial, to ensure that widespread adoption of the national contact-tracing app was not hampered by local and regional initiatives that had been in use for weeks before the national app was launched.

Involvement of Health Care Professionals

Designing a successful digital health solution requires advanced alignment of the interests of all health care professionals and stakeholders involved. Management of the COVID-19 pandemic is challenging for the health care systems in many ways, including maintaining a balance and relative responsibilities of hospital vis-à-vis community care. It has become clear over time that primary care professionals should play a central role in managing less severe COVID-19 cases via home-based monitoring, possibly by using digital solutions. However, in Italy, general practitioners are contracted professionals. Therefore, it has been much more complex to extend institutionally sponsored digital tools to these third parties and guarantee the necessary system interoperability between services deployed in primary care clinics and all others provided by LHAs. Elsewhere, however, this has been pivotal in implementing a successful digital strategy during the pandemic [7].

Coexistence of Digital and Analog Pathways for Health Care

The analog approach is not going away any time soon, so analog and digital health systems will need to coexist. Although digital health applications have spread in parallel to the virus, they have rarely been proposed as the only available option. None of the digital solutions implemented in the Italian SSN have been imposed as a mandatory alternative; rather, they have been added to the mix of traditional in-person and telephone-based services to maximize coverage for the entire population. Despite the increased adoption of smartphone use, Italy's overall population and health care workforce are among the "oldest" in the world [32]. The digital divide is still an issue, and COVID-19 has taught us that digital health is an important option to have in place, but it is no panacea.

Building on the Momentum

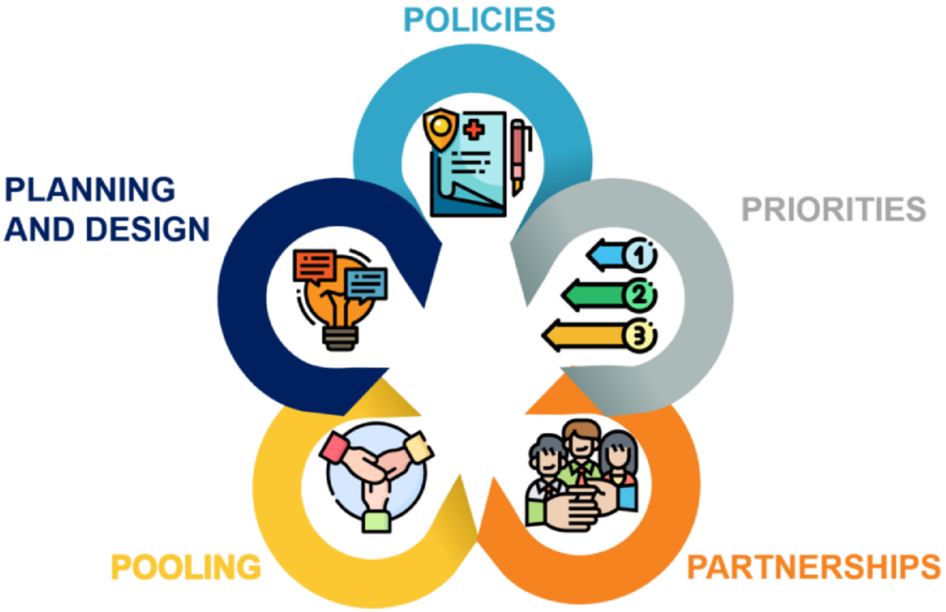
Despite the unprecedented expansion in the utilization of digital health tools during the COVID-19 pandemic, Italy, like many other countries worldwide, was not sufficiently equipped to harness the full potential of these tools. The urgency of the current public health emergency may have set the grounds for making exceptions [33], but the acceleration could not bypass the institutional and organizational elements that characterize national health care systems.

If we intend to maintain the momentum in expanding the digital health services that have resulted from the COVID-19 pandemic, now is the time to effectively plan for the future and to follow a new path, distinct from the pre-pandemic models [33]. The path must be selected by paying particular attention to avoid disparities in access to care. In theory, telemedicine can mitigate this issue but, if not carefully planned, it could also exacerbate it [34].

Achieving Sustained Benefits

No single best recipe exists, but a list of ingredients may help Italy, and other health care systems worldwide, focus their proposals to achieve sustained benefits (Figure 1):

Figure 1. Ingredients to achieve sustained benefits through digital health.



Policies

The COVID-19 crisis has reinforced the need to define and uphold appropriate policy standards. Long-standing digital health issues (reimbursement, regulatory framework, evidence generation, privacy, and security concerns) will still be unresolved by the end of this pandemic. Decisive and comprehensive action is thus needed from a policy standpoint, and several extensive proposals have been suggested in other contributions [1,4,18,23,33,35].

Priorities

The combination of the widespread availability of digital health solutions, their poor uptake before the pandemic, and the confidence gained during the recent expansion may push countries to aim too high. Governments should act promptly, but pragmatically, in adopting a stepped-wedge approach based on their health care system needs and organizational characteristics.

Partnerships

Digital transformation is bringing in new players with different backgrounds, expertise, and logic compared to the players that have typically populated the health care industry. Public institutions need to welcome innovation and be open to interorganizational relationships and trust. During a public health emergency, occasional, hastily generated partnerships may proliferate. Governments must adopt a strategic approach to partnerships and actively pursue the necessary competencies to harness these opportunities.

Pooling

The pre-pandemic digital health environment was characterized by a predominance of scattered experiences and the inability to adequately value existing best practices. Higher interinstitutional coordination should be assured by stewardship at the central level to take charge of streamlining the process, identifying parameters and conditions conducive to transferability, and ensuring no one is left behind—all without stifling local innovation.

Planning and Design

Digital health introduction should be complemented by an in-depth service redesign. The organizational implications of digital health have been largely neglected to date, but they need to be addressed now to exploit the complete potential of digital tools, integrate solutions in current care pathways, and pave the way for new models of care. Any digital technology is only as good as the response it gets from its end-users in terms of acceptability, continuity, and engagement. Hence, new, digitally enriched organizational processes should be planned and designed by all relevant stakeholders, directing patient-provider interactions to the channels generating the most value [36]. Irrespective of the cultural or contract issues, all health care professionals should be involved to steer change processes. Likewise, health care systems should flexibly partner with patients by enacting a mutual learning space and blurring the boundary between producers and consumers [37]. This shared effort at the system level should be pursued in the awareness that digital health literacy, among both citizens and health care professionals, still carries significant weight in influencing the uptake of digital solutions [38]. Although health care systems tackle this issue to grant the benefits of technological advancements to all users, the service redesign should be oriented by process of segmentation that will ideally result in a more tailored response to each individual’s needs through a combination of traditional and digital solutions.

Unprecedented times call for unprecedented decisions. It is time for policymakers to step up and make timely valiant choices about the use of telemedicine and digital health to permanently integrate them into the health care systems for the good of the public health and, ultimately, of the people. This will be extremely helpful now, during the COVID-19 crisis, and even more as we head toward the post-pandemic world.

Acknowledgments

We would like to warmly thank Ettore Turra for his contribution regarding the digital health solutions implemented in APSS Trento before and during the COVID-19 outbreak. The authors received no specific funding for this work.

Conflicts of Interest

FP, OC, MC, and RT all reported grants from the European Union’s Horizon 2020 research and innovation program under grant agreement no. 779306. FP, OC, MC, and RT are also involved in an RCT to evaluate a mobile supportive care app for patients with metastatic lung cancer.
Horizon 2020 research and innovation program under grant agreement no. 779306.

References

1. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
2. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
3. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global preparedness against COVID-19: we must leverage the power of digital health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
4. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
5. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
6. Lonergan PE, Washington Iii SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid utilization of telehealth in a comprehensive cancer center as a response to COVID-19: cross-sectional analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
7. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the crisis into an opportunity: digital health strategies deployed during the COVID-19 outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]
8. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine during the COVID-19 pandemic: experiences from western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
9. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
10. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
11. Magrabi F, Habli I, Sujan M, Wong D, Thimbleby H, Baker M, et al. Why is it so difficult to govern mobile apps in healthcare? BMJ Health Care Inform 2019 Nov;26(1):e100006 [FREE Full text] [doi: 10.1136/bmjhci-2019-100006] [Medline: 31744843]
12. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? Lancet 2020 Apr 11;395(10231):1225-1228 [FREE Full text] [doi: 10.1016/S0140-6736(20)30627-9] [Medline: 32178769]

13. Sorveglianza Integrata COVID-19 in Italia. L'epidemiologia per la sanità pubblica - Istituto Superiore di Sanità. 2020 Jun 22. URL: https://www.epicentro.iss.it/coronavirus/bollettino/Infografica_22giugno%20ITA.pdf [accessed 2020-06-24]
14. Coronavirus (COVID-19) Cases. Our World in Data. 2020 May 25. URL: https://ourworldindata.org/covid-cases [accessed 2020-06-24]
15. OECD and European Observatory on Health Systems and Policies. Italy: Country Health Profile 2019, State of Health in the EU. Brussels: OECD Publishing, Paris/European Observatory on Health Systems and Policies; 2019.
16. Rosenbaum L. Facing Covid-19 in Italy - ethics, logistics, and therapeutics on the epidemic's front line. N Engl J Med 2020 May 14;382(20):1873-1875. [doi: 10.1056/NEJMp2005492] [Medline: 32187459]
17. Paterlini M. On the front lines of coronavirus: the Italian response to covid-19. BMJ 2020 Mar 16;368:m1065. [doi: 10.1136/bmj.m1065] [Medline: 32179517]
18. Omboni S. Telemedicine during the COVID-19 in Italy: a missed opportunity? Telemed J E Health 2020 Aug;26(8):973-975 [FREE Full text] [doi: 10.1089/tmj.2020.0106] [Medline: 32324109]
19. Gabbrielli F, Bertinato L, De FG, Bonomini M, Cipolla M. Indicazioni ad interim per servizi assistenziali di telemedicina durante l'emergenza sanitaria COVID-19. Versione del 13 aprile 2020. Roma: Istituto Superiore di Sanità (Rapporto ISS COVID-19, n. 12/2020); 2020:29.
20. Multiple signatories. Contact Tracing Joint Statement. Research Group COSIC, Ku Leuven. 2020 Apr 19. URL: http://www.esat.kuleuven.be/cosic/sites/contact-tracing-joint-statement/ [accessed 2020-07-17]
21. Joint European Roadmap towards lifting COVID-19 containment measures. European Commission. 2020 Apr 08. URL: https://ec.europa.eu/info/sites/info/files/communication_-_a_european_roadmap_to_lifting_coronavirus_containment_measures_0.pdf [accessed 2020-06-02]
22. Apple and Google partner on COVID-19 contact tracing technology. Google Inc. Official Blog - Company Announcements. 2020 Apr 10. URL: https://www.blog.google/inside-google/company-announcements/apple-and-google-partner-covid-19-contact-tracing-technology/ [accessed 2020-12-23]
23. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 contact tracing and data protection can go together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
24. Cohen IG, Gostin LO, Weitzner DJ. Digital smartphone tracking for COVID-19: public health and civil liberties in tension. JAMA 2020 Jun 16;323(23):2371-2372. [doi: 10.1001/jama.2020.8570] [Medline: 32459289]
25. Immuni. Presidenza del Consiglio dei Ministri. 2020. URL: https://www.immuni.italia.it [accessed 2020-07-18]
26. App Immuni, la ministra Pisano: Scaricata da 4 milioni e 300 mila. Corriere della Sera Tecnologia. 2020 Jul 23. URL: https://www.corriere.it/tecnologia/20_luglio_23/app-immuni-ministra-pisano-scaricata-4-milioni-300-mila-3b39e218-ccf0-11ea-83db-f973956fabb4.shtml [accessed 2020-07-27]
27. Galligioni E, Piras EM, Galvagni M, Eccher C, Caramatti S, Zanolli D, et al. Integrating mHealth in oncology: experience in the province of Trento. J Med Internet Res 2015 May 13;17(5):e114 [FREE Full text] [doi: 10.2196/jmir.3743] [Medline: 25972226]
28. Eccher C, Botteri M, Ortolani D, Forti S, Piras EM. A mobile logbook to diagnose masked hypertension: a pilot application. Stud Health Technol Inform 2014;205:363-367. [Medline: 25160207]
29. Passardi A, Rizzo M, Maines F, Tondini C, Zambelli A, Vespignani R, et al. Optimisation and validation of a remote monitoring system (Onco-TreC) for home-based management of oral anticancer therapies: an Italian multicentre feasibility study. BMJ Open 2017 May 29;7(5):e014617 [FREE Full text] [doi: 10.1136/bmjopen-2016-014617] [Medline: 28554917]
30. #Accasa: il monitoraggio del virus a domicilio diventa digitale la Regione Puglia vara la piattaforma di teleassistenza clinica. Press Regione Regione Puglia. 2020 Apr 22. URL: https://press.regione.puglia.it/-/-accasa-il-monitoraggio-del-virus-a-domicilio-diventa-digitale-la-regione-puglia-vara-la-piattaforma-di-teleassistenza-clinica [accessed 2020-06-02]
31. TraceTogether. Government of Singapore. URL: https://www.tracetogether.gov.sg/ [accessed 2020-12-23]
32. Healthcare personnel statistics - physicians. Eurostat Statistics Explained. 2020 Nov. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_personnel_statistics_-_physicians#Healthcare_personnel [accessed 2020-06-05]
33. Shachar C, Engel J, Elwyn G. Implications for telehealth in a postpandemic future: regulatory and privacy issues. JAMA 2020 Jun 16;323(23):2375-2376. [doi: 10.1001/jama.2020.7943] [Medline: 32421170]
34. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst 2020 May 04:04 [FREE Full text]
35. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Sci Adv 2020 Jun;6(23):eabc0764 [FREE Full text] [doi: 10.1126/sciadv.abc0764] [Medline: 32548274]
36. Desruisseaux M, Stamenova V, Bhatia RS, Bhattacharyya O. Channel management in virtual care. NPJ Digit Med 2020;3:44 [FREE Full text] [doi: 10.1038/s41746-020-0252-4] [Medline: 32219185]
37. Batalden M, Batalden P, Margolis P, Seid M, Armstrong G, Opipari-Arrigan L, et al. Coproduction of healthcare service. BMJ Qual Saf 2016 Jul;25(7):509-517 [FREE Full text] [doi: 10.1136/bmjqs-2015-004315] [Medline: 26376674]
38. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]

Abbreviations

LHA: local health authority
SSN: Italian National Health Service

Edited by G Eysenbach, R Kukafka; submitted 25.06.20; peer-reviewed by J Piera-Jiménez, L Tizek, M Bardus, J Li; comments to author 13.07.20; revised version received 29.07.20; accepted 30.10.20; published 30.12.20.

Please cite as:
Petracca F, Ciani O, Cucciniello M, Tarricone R
Harnessing Digital Health Technologies During and After the COVID-19 Pandemic: Context Matters
J Med Internet Res 2020;22(12):e21815
URL: http://www.jmir.org/2020/12/e21815/
doi:10.2196/21815
PMID:33351777

Original Paper

Knowledge, Perceived Beliefs, and Preventive Behaviors Related to COVID-19 Among Chinese Older Adults: Cross-Sectional Web-Based Survey

Ying Chen[1*], PhD; Rui Zhou[1*], PhD; Boyan Chen[1], MSc; Hao Chen[1], MSc; Ying Li[2], MSc; Zhi Chen[3], PhD; Haihong Zhu[3], PhD; Hongmei Wang[1], PhD

[1]Department of Social Medicine of School of Public Health, and Department of Pharmacy of the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, China

[2]Department of Public Health, Xi’an Medical University, Xi’an, Shaanxi Province, China

[3]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, and Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, China

[*]these authors contributed equally

Corresponding Author:
Hongmei Wang, PhD
Department of Social Medicine of School of Public Health, and Department of Pharmacy of the First Affiliated Hospital
Zhejiang University School of Medicine
866 Yuhangtang Road
Xihu District
Hangzhou, Zhejiang Province, 310058
China
Phone: 86 571 88208222
Email: rosa@zju.edu.cn

Abstract

Background: The COVID-19 pandemic continues to pose an international public health threat. Prevention is of paramount importance to protect the high-risk group of older adults until specific treatments for COVID-19 become available; however, little work has been done to explore factors that promote preventive behaviors among this population.

Objective: This study aims to investigate the knowledge, perceived beliefs, and preventive behaviors towards COVID-19 of older adults in China and determine the factors that influence their practice of preventive behaviors.

Methods: From February 19 to March 19, 2020, a cross-sectional, web-based survey was administered to Chinese older adults in all 31 provinces of mainland China using a convenience sampling method to assess the respondents’ knowledge, perceived beliefs, and preventive behaviors towards COVID-19. Standard descriptive statistics and hierarchical linear regression analyses were conducted to analyze the data.

Results: A total of 1501 participants responded to the survey, and 1263 valid responses (84.1%) were obtained for further analysis. The overall correct rate on the knowledge questionnaire was 87%, overall positive beliefs regarding COVID-19 were found, and the mean behavior score was 13.73/15 (SD 1.62, range 5-15). The hierarchical linear regression showed that respondents who were married or cohabitating and who lived in areas with community-level control measures were more likely to practice preventive behaviors ($P<.01$). Knowledge ($\beta=0.198$, $P<.001$), perceived susceptibility ($\beta=0.263$, $P=.03$), perceived benefits ($\beta=0.643$, $P<.001$), and self-efficacy in preventing COVID-19 ($\beta=0.468$, $P<.001$) were also found to be significantly associated with preventive behaviors.

Conclusions: Most older residents had adequate knowledge and positive beliefs regarding COVID-19 and engaged in proactive behaviors to prevent the disease. Knowledge and beliefs were confirmed to be significantly associated with behavior responses. Our findings have significant implications in enhancing the effectiveness of COVID-19 prevention programs targeting the older population; these programs must be continued and strengthened as the epidemic continues.

(J Med Internet Res 2020;22(12):e23729) doi:10.2196/23729

KEYWORDS

COVID-19; knowledge; perceived beliefs; behaviors; elderly; China

Introduction

COVID-19, which was first detected in Wuhan, China, in December 2019, is an emerging infectious disease caused by a novel coronavirus (SARS-CoV-2). The crude case fatality rate of COVID-19 was 3.7% worldwide as of August 9, 2020 [1], which is much lower than that of severe acute respiratory syndrome (SARS, 9.5%), Middle East respiratory syndrome (MERS, 34.4%), or H7N9 (39.0%) [2]; however, evidence has shown that SARS-CoV-2 is more contagious than the viruses that cause the latter diseases [3,4]. Considering its extensive spread scope, COVID-19 was declared a public health emergency of international concern on January 30, 2020 [5] and was declared a pandemic on March 11, 2020 [6] by the World Health Organization (WHO). As of August 9, 2020, the virus had reached 215 countries and regions, resulting in over 19 million confirmed infections and more than 700,000 deaths [1].

Older people are especially susceptible to infectious diseases due to decreased immunity [7], the presence of underlying chronic illnesses [8], and cognitive impairment, which causes difficulty in their participating in proper prevention activities [9]. In the case of COVID-19, older people are also at significantly higher risk of morbidity and mortality [10-12]. According to an epidemiological study on COVID-19 in China, people aged >60 years comprised 31.2% of the 44,672 confirmed cases and 81% of the total number of deaths [13].

To date, no specific treatments or vaccines for COVID-19 have been made available; thus, it is of great importance for older people to engage in preventive behaviors, which has been proved to be effective in preventing respiratory infectious diseases and controlling their transmission [14,15]. It is believed that the transmission route of COVID-19 is human-to-human via respiratory droplets or direct contacts [16] and that the average infected person can spread the disease to up to 3 other people [17]; thus, effective preventive measures, including handwashing, mask-wearing, household ventilation and disinfection, and reduction of interpersonal contacts by avoiding visiting crowded spaces [18], have been recommended to older people by local Chinese authorities. However, these guidelines would severely hamper many daily activities if implemented at a high level of fidelity; therefore, voluntary compliance is likely to be uneven at best.

In fact, in the early stage of the COVID-19 epidemic, older people underestimated the severity of the epidemic, as it affected a relatively small population. They were reluctant to listen to the authorities, believing that there was no need to practice the suggested preventive behaviors for COVID-19. A recent study found that compared with younger adults, older men were less worried about COVID-19 and adopted fewer preventive behaviors [19]. Hence, special efforts should be made to enhance older people's protection against this new infectious disease.

A good understanding of underlying factors that encourage older people to adopt particular preventive behaviors is significant to provide necessary strategies to implement. A higher level of knowledge has been proved to be associated with positive behavior changes [20,21]. In addition, studies on individuals' behavior changes toward infectious diseases have suggested that perceptions or beliefs about an outbreak are important in determining the adoption of particular preventive behaviors [22-24]. In particular, the health belief model (HBM) [25] has provided a theoretical framework in previous research for understanding and explaining various health-promoting behaviors [26-28], including preventive behaviors during global pandemics [29-31]. The HBM model reveals that significant perception factors include perceived severity, perceived susceptibility, and perceived benefits. Additionally, previous literature has suggested that self-efficacy in preventing disease [32] and belief in the ability of the government to contain the spread of the epidemic [33] are significantly associated with the practice of preventive measures.

Studies have investigated behavioral responses toward COVID-19 among the general population [34,35] and in specific populations such as health care workers [36] and college students [37,38]; however, literature reports focusing on behavior changes in older people remain scant in spite of the detection of increased vulnerability among this population. Therefore, this paper aims to assess the knowledge, perceived beliefs, and preventive behaviors of older Chinese residents toward COVID-19 and to explore the knowledge, perceived beliefs, and other relevant factors that influence behavior changes. Our findings may help enhance the effectiveness of intervention programs targeting the older population.

Methods

Study Design

This cross-sectional study was conducted using an anonymous web-based questionnaire between February 19 and March 19, 2020. A convenience sampling method was used to recruit participants. Chinese residents aged ≥60 years without intellectual or cognitive impairment who agreed to participate in our survey were considered eligible. Those who were illiterate or unable to use electronic devices could ask others to help them fill out the questionnaire; however, it was emphasized that the answers must always reflect the older person's own opinions. The link to the web-based questionnaire was sent through WeChat (which is similar to WhatsApp) and other social media platforms to the contacts of potential respondents. A brief introduction was presented at the beginning of the survey to inform the respondents of the purpose and content of this study and to instruct them on how to complete the questionnaire. Respondents who completed the questionnaire were entered in a draw for a monetary incentive of approximately ¥1-5 (US $0.15-0.76).

Sampling

According to Kendall, multiple regression analysis demands a sample size of at least 5 to 10 times the number of independent variables. Thus, we required a maximum sample size of 670 cases, as 67 independent variables were included in this study. Returned surveys with unreasonably short answer times, incomplete information, or logic problems were deemed invalid. The effective sample size estimated for the study was 957 cases, with an invalid response rate of 30%. Finally, a total of 1501 questionnaires were returned in the study duration; after excluding 212 invalid questionnaires and 26 participants who reported having never heard about COVID-19, a total of 1263 surveys (84.1%) from all of the 31 provinces, municipalities, and autonomous regions in mainland China remained for statistical analysis.

Measures

Dependent Variable (Practice of Preventive Behaviors)

The questionnaire on the practice of COVID-19 preventive behaviors was developed based on the guidelines from the Chinese Center for Disease Control and Prevention (CDC). This section contained 5 items, including washing hands frequently, wearing a face mask in public areas, disinfecting one's household, ventilating one's household, and avoiding crowds (P1-P5). Answers of “more,” “as usual,” and “less” were rated as 3, 2, and 1, respectively. Therefore, the total behavior score could range from 5 to 15, with a higher score indicating better adherence to preventive behaviors.

Independent Variables

Background Variables

Sociodemographic characteristics were surveyed in this section of the survey, including gender, age, marital status, education level, registered residential area, and monthly household income per capita. Additionally, the respondents' self-perceived health status, current addresses at the district level, and local community-level control measures (free entry/exit as usual, entry/exit control exercised, and lockdown) in their residential areas were surveyed. The provinces of the respondents' current addresses were divided into eastern, central, and western regions according to the National Bureau of Statistics of China and were also categorized into three levels according to their numbers of confirmed COVID-19 cases; provinces with <100 cumulative confirmed cases, 100-1000 cases, and >1000 cases were rated as low, medium, and high-risk areas, respectively.

COVID-19–Related Knowledge

The participants' levels of knowledge about COVID-19 were assessed by 7 questions (K1-K7): 1 regarding the province in which the first COVID-19 cases were reported, 1 regarding the incubation period, 1 regarding the source of infection, 2 regarding transmission modes, 1 regarding susceptible populations, and 1 regarding the availability of a vaccine for COVID-19. The answers were judged according to the guidelines for clinical management of COVID-19 by the National Health Commission of the People's Republic of China [39]. A correct answer was assigned 1 point, and an incorrect or unknown answer was assigned 0 points. The total knowledge score thus ranged from 0 to 7, with a higher score representing better knowledge of COVID-19.

Perceived Beliefs Regarding COVID-19

Based on the HBM and previous studies, the perceived beliefs measured in our study included constructs of HBM (perceived severity, perceived susceptibility, and perceived benefits), self-efficacy in preventing the disease, and belief that the government can contain the spread of the epidemic (B1-B5). Perceived severity of COVID-19 was assessed by 1 item asking the respondents whether they would suffer from more severe symptoms if they contracted the disease; an answer of “yes” was coded as 1, while other answers were coded as 0. One item assessed perceived susceptibility by measuring the degree to which the respondents perceived themselves to be vulnerable to COVID-19; the responses were coded into two groups (1: high/very high; 0: low/very low/I don't know). Perceived benefits were assessed by three items measuring the respondent's belief in the effectiveness of COVID-19–related preventive measures of mask wearing in public areas, handwashing, and avoidance of visiting crowded places, respectively; answers of “yes” were considered appropriate (1: yes; 0: others), and the perceived benefits indicator was established by counting the number of appropriate answers (values of 0-3). The respondents' self-efficacy in preventing the disease and confidence in the government to contain the spread of the epidemic were assessed by asking “Are you confident that you can protect yourself from contracting COVID-19?” and “Do you believe that the government can win the battle against COVID-19?” Answers of “yes” were considered positive and were given 1 point, while 0 points were given for choosing other answers.

Statistical Analysis

Descriptive statistics were used to summarize the background factors and COVID-19 knowledge, perceived beliefs, and behaviors; the results are presented as frequencies (n) and percentages (%) or as means and SDs. Associations between background variables and behavior scores were examined by one-way analysis of variance (ANOVA) or independent sample *t* tests as appropriate. A multiple linear regression model using statistically significant background variables, knowledge scores, and factors of perceived beliefs as independent variables and scores of preventive behavior as outcome variables was applied to identify factors associated with preventive behaviors. The demographic variables of the respondents were entered in the regression model first to control for their effects. Then, the knowledge score was entered in the next block of the regression, and factors of perceived beliefs were simultaneously entered in the last block. Unstandardized regression coefficients (β) and 95% CIs are reported. All statistical analyses were performed in SPSS version 21.0 (IBM Corporation), and a *P* value <.05 was considered statistically significant.

Ethics Statement

The study protocol was approved by the Ethics Committee of School of Public Health, Zhejiang University (approval number: ZGL202002-2) before the formal survey. The questionnaire was designed to be anonymous and voluntary, and respondents were informed that submission of the questionnaire implied informed consent. The data were kept confidential, and the results did not identify the respondents personally.

Results

Background Characteristics

Of the 1263 participants, 730 (57.8%) completed the questionnaire with the help of others, and 687 (54.4%) were registered permanent residents of rural areas. The mean age was 69.48 years (SD 6.72), over half of the respondents were female (697/1263, 55.2%), and approximately three-fourths were married or cohabitating (941/1263, 74.5%). Of the 1263 participants, 586 (46.4%) had an education level of primary school or below, and most respondents (844, 66.8%) had an average household income of between ¥600 and ¥6000 per month (US $91.61-$916.07). Moreover, 564/1263 participants (44.7%) self-reported their physical health status to be fair; 613 (48.5%) of the 1263 respondents lived in the eastern region, and nearly half of the participants (591/1263, 46.8%) lived in areas with medium risk of COVID-19. The majority of respondents (1050/1263, 83.1%) reported that entry and exit control was exercised in their community or village.

Levels of COVID-19–Related Knowledge, Perceived Beliefs, and Behaviors

Knowledge

The mean knowledge score was 6.06 (SD 0.03, range 1-7), suggesting an overall 87% (6.06/7*100) correct rate on this knowledge test (Table 1). Of the 1263 respondents, 85.4% (n=1078) agreed that the current main source of infection was patients with COVID-19. In terms of modes of transmission, 98.1% of respondents (1239/1263) thought that COVID-19 can be transmitted by droplets emitted by patients; however, less than 80% (960/1263, 76.0%) were aware that it can also be transmitted through virus-contaminated objects. Approximately 90% of respondents (1165/1263, 92.2%) correctly stated that the population is generally susceptible to the virus regardless of age. However, only 64.1% of the respondents (809/1263) knew that effective vaccines for COVID-19 were not yet available.

Table 1. Participants' knowledge and perceived beliefs regarding COVID-19 (N=1263).

Item	Response
Knowledge items	
K1: The earliest outbreak of COVID-19 in China occurred in Hubei Province, n (%)	
Correct	1258 (99.6)
Incorrect/unknown	5 (0.4)
K2: The incubation period of COVID-19 is 1-14 days, n (%)	
Correct	1139 (90.2)
Incorrect/unknown	124 (9.8)
K3: Currently, the main source of infection is patients with COVID-19, n (%)	
Correct	1078 (85.4)
Incorrect/unknown	185 (14.6)
K4: COVID-19 can be transmitted by droplets from a patient, n (%)	
Correct	1239 (98.1)
Incorrect/unknown	24 (1.9)
K5: COVID-19 can be transmitted through touching virus-contaminated surfaces, n (%)	
Correct	960 (76.0)
Incorrect/unknown	303 (24.0)
K6: All age groups can become infected with the new coronavirus, n (%)	
Correct	1165 (92.2)
Incorrect/unknown	98 (7.8)
K7: There is currently no vaccine available that protects against COVID-19, n (%)	
Correct	809 (64.1)
Incorrect/unknown	454 (35.9)
Range of knowledge scores	1-7
Mean knowledge score (SD)	6.06 (0.03)
Perceived belief items, n (%)	
B1: Perceived severity of COVID-19	
Yes	1141 (90.3)
No/I don't know	122 (9.7)
B2: Perceived susceptibility of COVID-19	
High/very high	190(15.0)
Low/very low/I don't know	1073(85.0)
B3: Perceived benefits indicator[a]	
0	6 (0.5)
1	26 (2.1)
2	68 (5.4)
3	1163 (92.1)
B4: Self-efficacy in preventing COVID-19	
Yes	1048 (83.0)
No/I don't know	215 (17.0)
B5: Confidence in the government to control the spread of COVID-19	
Yes	1160 (91.8)
No/I don't know	103 (8.2)

[a] The perceived benefits indicator was established by counting the number of items with “yes” responses for the three basic preventive measures (ie, handwashing, face mask wearing, and staying at home).

Perceived Beliefs

The majority of the older respondents (1141/1263, 90.3%) believed that older people may suffer from more severe symptoms if they are infected with COVID-19 (Table 1). Most respondents (1073/1263, 85.0%) did not perceive that their likelihood of contracting COVID-19 was high or very high. A substantial proportion of participants (1163/1263, 92.1%) perceived the benefits of all three types of preventive measures (ie, handwashing, wearing a face mask, and staying at home). Additionally, 83.0% of the respondents (1048/1263) were confident in their own abilities to prevent the disease, and 91.8% (1160/1263) believed that the government can “win the battle” against COVID-19.

Behaviors

The mean score of preventive behavior toward COVID-19 was 13.73 (SD 1.62, range 7-15). The participants’ preventive behavior changes in response to the COVID-19 pandemic are shown in Table 2. It is reassuring that more than four-fifths of our 1263 participants had increased their frequency of handwashing (n=1025, 81.2%) and wearing a face mask in public venues (n=1035, 81.9%) to ensure their safety; in addition, 86.6% (n=1094) of the older people had reduced their visits to crowded places. However, fewer participants had increased their household ventilation (956/1263, 75.5%) and disinfection (816/1263, 64.6%).

Table 2. Participants’ preventive behavior changes in response to the COVID-19 pandemic in the past week (N=1263), n (%). Questions were answered using the following scale: 1=less, 2=as usual, and 3=more. For the total behavior score, we calculated how many specific behavior changes each participant endorsed. The range of the behavior scores was 7-15, and the mean score was 13.73 (SD 1.62).

Behavior	More	Less	As usual
P1: Face mask wearing in public venues	1035 (81.9)	60 (4.8)	168 (13.3)
P2: Washing of hands	1025 (81.2)	12 (1.0)	226 (17.9)
P3: Household ventilation	956 (75.5)	66 (5.2)	241 (19.1)
P4: Home disinfection	816 (64.6)	38 (3.0)	409 (32.4)
P5: Avoidance of crowds	1094 (86.6)	37 (2.9)	132 (10.5)

Demographic Factors of Preventive Behaviors: Univariate Analysis

As shown in Table 3, region, marital status, education level, residence registration, and monthly household income per capita were significantly associated with preventive behaviors (all *P*<.05). Participants who lived in western regions had significantly lower level of behavior than participants from eastern regions (*P*=.01). Respondents who were married or cohabitating had significantly higher behavior scores (*P*<.001). Rural dwellers and respondents who attended primary school or below had lower behavior scores (both *P*<.001). The behavior scores of participants with an average household income of <¥600 per month (US $91.61) were significantly lower than those of participants with monthly incomes higher than ¥6000 (US $916.07; *P*=.02). Interestingly, a significant association between local community-level control measures and behavior score was observed (*P*<.001); respondents engaged in more preventive behaviors when epidemic control measures were exercised in their areas of residence.

Table 3. Univariate analysis of the demographic variables associated with preventive behaviors (N=1263).

Variable	Participants, n (%)	Behavior score[a]		
		Mean (SD)	*t*/*F* (df)	*P* value
Gender			t_{1262}=0.208	.65
Male	566 (44.8)	13.71 (1.63)		
Female	697 (55.2)	13.75 (1.61)		
Age (years)			$F_{(2,1260)}$=2.925	.054
60-69	683 (54.1)	13.81 (1.61)		
70-79	474 (37.5)	13.69 (1.62)		
≥80	106 (8.4)	13.42 (1.67)		
Marital status[b]			t_{1262}=18.625	*<.001*[c]
Married/cohabiting	941 (74.5)	13.85 (1.56)		
Single/divorced/ separated/widowed	322 (25.5)	13.40 (1.74)		
Education level[b]			$F_{(3,1259)}$=8.904	*<.001*
Primary education or below	586 (46.4)	13.50 (1.67)[d]		
Middle school	330 (26.1)	13.94 (1.57)[e]		
High school	197 (15.6)	13.79 (1.65)[e]		
College or above	150 (11.9)	14.11 (1.33)[e]		
Registered residential area[b]			t_{1262}=20.090	*<.001*
Urban	576 (45.6)	13.95 (1.52)		
Rural	687 (54.4)	13.55 (1.68)		
Monthly household income per capita (¥)[b,f]				
<600	219 (17.3)	13.51 (1.78)[d]	$F_{(2,1260)}$=4.031	*.02*
600-6000	844 (66.8)	13.74 (1.63)[d,e]		
>6000	200 (15.8)	13.96 (1.36)[e]		
Self-reported health status			$F_{(4,1258)}$=0.375	.83
Excellent	57 (4.5)	13.72 (1.68)		
Very good	220 (17.4)	13.85 (1.58)		
Good	378 (29.9)	13.68 (1.73)		
Fair	564 (44.7)	13.72 (1.57)		
Poor	44 (3.5)	13.75 (1.53)		
Region[b]			$F_{(2,1260)}$=4.393	*.01*
Eastern	613 (48.5)	13.86 (1.49)[d]		
Central	183 (14.5)	13.72 (1.64)[d,e]		
Western	467 (37.0)	13.57 (1.77)[e]		
Provincial COVID-19 epidemic level[g]			$F_{(2,1260)}$=1.166	.31
Low risk: <100 cases	146 (11.6)	13.73 (1.68)		
Medium-risk: 100-999 cases	591 (46.8)	13.66 (1.69)		
High risk: ≥1000 cases	526 (41.6)	13.81 (1.52)		
Local community-level control measures[b]			$F_{(2,1260)}$=7.634	*<.001*

Variable	Participants, n (%)	Behavior score[a] Mean (SD)	t/F (df)	P value
Entry/exit control exercised	1050 (83.1)	13.70 (1.63)[d]		
Lockdown	181 (14.3)	14.04 (1.45)[d]		
Free entry/exit as usual	32 (2.5)	12.91 (1.91)[e]		

[a] Values based on independent sample *t* tests and one-way analysis of variance (ANOVA) for continuous variables to examine differences between preventive behaviors and demographic variables. For categories of variables with significant ANOVA results (significant at *P*<.05 between the groups), multiple comparisons between each of 2 categories were performed by post hoc analysis (least significant difference).

[b] Significant variables were included in the subsequent multivariable analyses.

[c] Italics indicate statistical significance.

[d,e] Within each column, if two means share same superscript (d or e), they are not statistically different (*P*>.05) from one another.

[f] 1 ¥=US $0.14 on February 20, 2020.

[g] Provincial COVID-19 epidemic level: The number of cumulative confirmed COVID-19 cases on the survey launch day (ie, March 20, 2020) in the province where the respondent was located during the COVID-19 pandemic.

Factors Associated With Behaviors: Multivariable Analyses

The knowledge scores and perceived belief-related variables reported in Table 1, together with significant background variables, were included in the hierarchical linear regression models. In block 1, the demographic factors of region, marital status, education level, registered residence, monthly household income per capita, and local community-level control measures were entered first, accounting for 4.7% of the variance (R^2=0.047, $F_{6,1256}$=5.575, *P*<.001). The score for knowledge about COVID-19 was entered into block 2 and contributed an additional 2.9% of the variance (R^2=0.076, $F_{7,1255}$=8.603, *P*<.001). Finally, factors of perceived beliefs regarding COVID-19, including perceived severity, perceived susceptibility, perceived benefits, self-efficacy in preventing the disease, and confidence in the government, were entered in the third block; these factors were significant in predicting behaviors even after considering the effects of background and knowledge factors, and they contributed an additional 3.2% of the variance (R^2=0.108, $F_{12,1250}$= 8.904, *P*<.001). The linear regression analysis of the preventive behaviors is presented in Table 4 and Table 5.

Education level and registered residence were generally nonsignificant, while some significant marital status and control measure differences were noted. Overall, respondents who were married or cohabitating (vs single, divorced, separated, or widowed, β=0.355, 95% CI 0.150-0.560) and who could not leave the house (vs free entry and exit, β=0.898, 95% CI 0.310-1.486) were more likely to have better preventive behavior scores (*P*<.01). After adjusting for background characteristics, we discovered that older people with higher COVID-19–related knowledge scores had significantly higher behavior scores (β=0.198, *P*<.001). Meanwhile, the results showed that 3 of the 5 components of perceived beliefs, namely, perceived susceptibility (β=0.263, 95% CI 0.022-0.504), perceived benefits (β =0.643, 95% CI 0.305-0.982), and self-efficacy in preventing the disease (β =0.468, 95% CI 0.223-0.713), were positively associated with preventive behaviors (all *P*<.05); however, perceived severity and confidence in the government were not significant predictors of preventive behaviors. Among the influencing factors, local community-level control measures showed the greatest impact on behaviors, followed by perceived benefits.

Table 4. Multivariable analysis of factors associated with preventive behaviors (N=1263).

Characteristic	Model 1		Model 2		Model 3	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
Block 1: Background characteristics						
Region						
Western	Reference	N/A[a]	Reference	N/A	Reference	N/A
Eastern	0.156 (0.049 to 0.361)	.14	0.082 (–0.121 to 0.285)	.43	0.085 (–0.116 to 0.286)	.41
Central	0.087 (–0.186 to 0.360)	.53	0.013 (–0.256 to 0.283)	.92	–0.051 (–0.318 to –0.217)	.71
Marital status						
Single/divorced/separated/widowed	Reference	N/A	Reference	N/A	Reference	N/A
Married/cohabitating	0.330 (0.120 to 0.541)	*.002*[b]	0.331 (0.124 to 0.538)	*.002*	0.355 (0.150 to 0.560)	*<.001*
Education level						
Primary education or below	Reference	N/A	Reference	N/A	Reference	N/A
Middle school	0.244 (0.014 to 0.475)	*.04*	0.177 (–0.050 to 0.405)	.13	0.137 (–0.088 to 0.362)	.23
High school	–0.002 (–0.293 to 0.290)	.99	–0.106 (–0.396 to 0.183)	.47	–0.117 (–0.402 to 0.168)	.42
College or above	0.258 (–0.078 to 0.594)	.13	0.123 (–0.210 to 0.457)	.47	0.096 (–0.233 to 0.425)	.57
Registered residential area						
Rural	Reference	N/A	Reference	N/A	Reference	N/A
Urban	0.243 (0.031 to 0.455)	*.02*	0.184 (0.025 to 0.393)	.09	0.122 (0.085 to 0.329)	.25
Monthly household income per capita (¥)[c]						
<600	Reference	N/A	Reference	N/A	Reference	N/A
600-6000	0.047 (–0.203 to 0.297)	.71	0.031 (–0.215 to 0.422)	.81	0.018 (–0.226 to 0.261)	.89
>6000	0.113 (–0.229 to 0.454)	.52	0.086 (–0.251 to 0.422)	.62	0.067 (–0.266 to 0.399)	.70
Local community-level control measures						
Free entry/exit as usual	Reference	N/A	Reference	N/A	Reference	N/A
Entry/exit control exercised	0.671 (0.108 to 1.234)	*.02*	0.635 (0.081 to 1.190)	*.03*	0.501 (–0.049, 1.051)	.07
Lockdown	1.062 (0.460 to 1.663)	*<.001*	1.034 (0.441 to 1.627)	*<.001*	0.898 (0.310 to 1.486)	*.003*
Block 2: Knowledge						
Knowledge related to COVID-19	—[d]	—	0.270 (0.186 to 0.353)	*<.001*	0.198 (0.111 to 0.286)	*<.001*
Block 3: Perceived beliefs						
Perceived severity	—	—	—	—	–0.045 (–0.350 to 0.259)	.77
Perceived susceptibility	—	—	—	—	0.263 (0.022 to 0.504)	*.03*
Perceived benefits indicator[e]	—	—	—	—	0.643 (0.305 to 0.982)	*<.001*
Self-efficacy in preventing COVID-19	—	—	—	—	0.468 (0.223 to 0.713)	*<.001*
Confidence in the government to fight COVID-19	—	—	—	—	0.321 (–0.014 to 0.656)	.06

[a] N/A: not applicable.

[b] Italics indicate statistical significance between groups.

[c] 1 ¥=US $0.14 on February 20, 2020.

[d] —: Data not included in this model.

[e] The perceived benefits indicator was established by counting the number of items with "yes" responses for the three basic preventive measures (ie,

handwashing, face mask wearing, and staying at home).

Table 5. Statistical measures of the 3 models of the multivariable analysis (*P*<.001 for all models).

Measure	Model 1	Model 2	Model 3
F	5.575	8.603	8.904
R^2	.047	.076	.108
ΔR^2	.047	.029	.032
Adjusted R^2	.038	.067	.096

Discussion

Principal Findings

Older people, who are particularly vulnerable to acute diseases and their complications, are currently greatly threatened by the outbreak of COVID-19. Special effort should therefore be made to encourage older people to practice the behaviors suggested by the government to prevent the disease. An understanding of older people's knowledge, perceived beliefs, and behaviors toward COVID-19, which affect the adoption of related health behaviors, can be a first step to prevent the spread of COVID-19 in this population.

In this study, it was found that Chinese older adults had good knowledge of COVID-19; the respondents had an overall correct rate of 87% on the knowledge questionnaire, which was slightly lower than a rate reported for the Chinese general population (90%) [40] but much higher than that reported for US residents (80%) [41]. The high correct rate was surprising, as previous studies suggested that knowledge scores are negatively related with age [42-44]. However, this finding may be due to the large amounts of publicity related to COVID-19 through various channels that are appropriate to the needs and characteristics of older people, such as vivid prints, marked banners, and broadcasts in dialect. The 2 items with the lowest correct rates are worth mentioning. Knowledge about modes of transmission, which has been proved to be a salient factor influencing the level of adoption of preventive measures [45], should be further improved among older people. Meanwhile, older people should also be informed that an effective vaccine for COVID-19 is not yet available and that taking precautions against the disease is still highly important.

The perceived beliefs of older Chinese residents toward COVID-19 were found to be optimistic overall. Although the majority of respondents (1141/1263, 90.3%) perceived that older adults would suffer from more severe symptoms if they were infected, relatively few respondents (190/1263, 15.0%) thought they were at high risk of acquiring the disease. Moreover, the majority expressed confidence in themselves (1048/1263, 83.0%) and the government (1160/1263, 91.8%) to stop the spread of COVID-19. This optimistic belief could be explained by the unprecedented epidemic control measures taken by the Chinese government after the outbreak of COVID-19 in the country and the concerted efforts of people all over China to prevent the spread of the disease. From January 20, 2020, when person-to-person transmission was confirmed and the Chinese public was notified of this finding, a series of nationwide public health emergency measures, including isolation and quarantine, close management of working and living spaces, and the Examine and Approve Policy on the resumption of work, were implemented by the Chinese government, health institutions, communities, companies, etc [46].

In addition to efforts from all sectors of society, high adherence to the lifestyle modifications suggested by the government and public health organizations greatly delayed the spread of the disease [47]. In our findings, 64.6%-86.6% of respondents reported practicing five major preventive behaviors more frequently in the preceding week. The reason for this finding may be that this study was conducted at a time when these key preventive measures were highly emphasized in China. Similarly high tendencies to adopt these precautionary behaviors during the pandemic have been found in other studies [35,48]. In particular, more than 80% of respondents reported increased handwashing and face mask wearing and decreased time spent in crowds, indicating that basic protective behaviors against COVID-19 had taken root among older people. The majority of participants avoided going out, which can potentially be attributed to the widely disseminated governmental propaganda and social media messages, which continuously instilled and reinforced people's incentives to stay home (eg, personal safety, good citizenship, and contribution to the control of the national epidemic). Moreover, individuals may have felt pressure from the community to adhere to behaviors such as imposed lockdowns or quarantine of entire regions with suspected cases. Handwashing and wearing of face masks have long been regarded as significant preventive habits in the daily life of the public since they were proved to be efficacious in preventing influenza and SARS [49,50]. Lau et al [51] observed that hand-washing and face mask wearing were commonly practiced during the SARS outbreak, and these behaviors were sustained by a large proportion of the public even after the SARS epidemic subsided in Hong Kong. In contrast, the effectiveness of household ventilation and disinfection may have been neglected, with fewer participants reporting that they had practiced these two behaviors more frequently in the past week. Habit has a great impact on routine behavior—including hygiene behavior—and despite their best intentions, people may find it difficult to implement new measures during a pandemic if they were not previously in the habit of performing them [52,53]. Further education on practicing household ventilation and disinfection is needed, as these two measures are indispensable in preventing COVID-19.

In line with previous research [54-56], our study also found that individuals' demographic characteristics had a significant influence on their preventive behaviors. Linear regression analysis of the behavior scores showed that participants who were married or cohabitating were more likely to report that they had complied with advocated protective behaviors more frequently; this finding was consistent with similar studies regarding SARS [57,58]. On the one hand, older adults who are married or cohabiting can receive support from their spouse or partner. Family support has been shown to influence older people's health beliefs and self-care behaviors [59]. The more family support older people receive, the more attention they will pay to their own health, accompanied with increased willingness to acquire health knowledge and more positive preventive beliefs. On the other hand, according to the HBM, advice from family members can be regarded as an external cue to action, which is also a very important factor in increasing various preventive behaviors [60]. Consistent with research by Kong et al [48], age was not a significant factor of preventive behaviors in our study (*P*=.054); meanwhile, other studies have found that compared with younger participants, individuals aged 60 years or older implemented fewer preventive behavior changes [19,61]. Further larger-scale studies are warranted to assess the association between age and behaviors to prevent COVID-19.

Additionally, community-level control measures were found to be significantly associated with the practice of preventive measures. As COVID-19 rapidly spread from a single city to the entire country, the governmental "minimum contact strategy" was implemented nationwide; also, close management and screening of communities and villages to curb COVID-19 were subsequently implemented nationally in China [46]. Although the contribution of community control measures cannot be quantified, our study supports that community-level control measures against COVID-19 are related to significant increases in older adults' practice of recommended preventive behaviors. We further confirmed that respondents who could not leave the house showed higher compliance with suggested preventive measures than those who could move freely as usual. According to the socioecological model, factors of structural, interpersonal, and personal levels determine health-related behaviors [62]. At the personal level, it has been found that older Chinese people feel they have an ethical duty ("filial piety") to protect others, which facilitates their adherence to quarantine. In Wenzel's study [63], "filial piety" suggested the "right and humane" way to act toward one's family and others in the community and guided Chinese older adults in their use of strategies against SARS [63]. As social beings, people's behaviors are subject to the influence of social relationships [64]. Chinese older people not only worry for themselves but also have concern for their significant others regarding COVID-19, which motivates their adoption of precautionary measures. Meanwhile, environmental manipulation and policies are crucial structural-level factors to encourage individuals to practice desired preventive behaviors. According to the WHO, social distancing/self-isolation and lockdown are two important nationwide social measures during a public health crisis [6]. The Chinese government adopted these strategies to mitigate the risk of COVID-19. Therefore, personal perceived ethical duty combined with community actions have facilitated China's response to COVID-19. Particularly, community and home quarantine provided Chinese older adults with ample time to communicate with and influence their significant others. The abovementioned findings highlight the importance of control measures at the community level, which should be considered when planning pandemic control strategies in the future.

In addition to background factors, this study found that knowledge and perceived beliefs toward COVID-19 were significant predictors of preventive behaviors. In accordance with previous findings, this study confirmed that people with adequate knowledge about COVID-19 were more likely to take preventive measures in response to the pandemic than people who lacked such knowledge [56,65]; this suggests that health education aimed at improving people's COVID-19 knowledge plays an important role in promoting preventive behaviors toward COVID-19. With regard to perceived belief–related factors, our study confirmed the positive correlation between self-efficacy in preventing COVID-19 and the adoption of preventive measures found in previous research [32,54]; however, bidirectional effects may exist between these two variables, as it has been suggested that perceived self-efficacy can encourage an individual to adopt certain behaviors to achieve desired outcomes, while success in performing certain behaviors can further augment perceived self-efficacy [66]. Contrary to previous findings [22,30,33], our study found that belief in the efficacy of local health authorities did not significantly predict the target COVID-19 preventive behaviors. The belief in government may not successfully translate into positive behavioral changes at the individual level; additionally, the positive relationship between trust in the government and the practice of preventive behaviors can be mediated by levels of perceived self-efficacy [20]. A relatively strong linear correlation between confidence in the government and self-efficacy (r=0.335, P<.001) was also found in this study.

Some of the items we used to assess the participants' perceived beliefs were derived from the HBM, which includes perceived severity, perceived susceptibility, and perceived benefits. Although most of our respondents perceived that they might progress to serious symptoms if they were infected with COVID-19, perceived severity was not found to be significantly related to COVID-19–related preventive behaviors in this study. Perceived severity has been suggested to have relatively low relevance for preventive health behaviors; however, it may play an important role when individuals have already been diagnosed with certain diseases [25]. Therefore, the fact that the participants had not yet been diagnosed with COVID-19 as well as their low perceived susceptibility could account for the insignificant association between perceived severity and preventive behaviors. Among our participants, perceived susceptibility was found to be a significant correlate of COVID-19 preventive health behaviors. Perceived susceptibility has been consistently found to be a salient determinant of participation in preventive measures during an epidemic among both the general population and older people [23,33,67,68]. Other studies on health screening [69,70] and exercise behavior [71] also found that perceived susceptibility plays an important role in participation in protective behaviors. Despite the significant positive association between perceived susceptibility and engagement in protective behaviors, only 15.0% (190/1263) of the older people in this study perceived that they were highly vulnerable to COVID-19. This finding indicates that older people should be further warned about their higher vulnerability to the disease; moreover, to ensure the effectiveness of warnings, messages issued by public health communicators should be comprehensible, concise, and convincing [22]. Meanwhile, perceived benefits were found to be more prominent in affecting individuals' adoption of preventive behaviors in the aforementioned HBM constructs [26]. When people deem certain preventive health behaviors to be effective in preventing a disease, they are motivated to engage in these behaviors [22,58,72,73]. Indeed, communications that increase individuals' perceptions of the benefits of particular health-related behaviors have been proved to be successful in reducing health threats [74-76]. Health authorities should be reminded to exert effort to provide sufficient health communication messages to enhance older people's perceived susceptibility and benefits in the context of the COVID-19 pandemic, which can greatly motivate these individuals to engage in preventive measures.

Limitations

Our study has some limitations. First, no standardized tool for assessing knowledge, perceived beliefs, or practice of preventive behaviors on COVID-19 has been previously validated. We designed the questionnaire based on the latest official report from the WHO, the Chinese CDC, and the scientific literature; however, the depth of the survey may be limited, as we were obligated to conduct our investigation quickly and recruit as large a sample as possible. Secondly, selection bias may exist. The web-based survey only included people who have access to the internet, and older people who were infected or had close contacts with confirmed or suspected COVID-19 patients may not have wanted to participate in this survey; thus, the generalizability of our findings may be limited. However, web-based surveys may be the most appropriate method of data collection during an epidemic, as this method can prevent transmission; moreover, fielding the survey offline was not feasible, as strict epidemic control measures were being exercised in most parts of the country. Thirdly, this study was cross-sectional, and no casual effect statements concerning the relationship between targeting variables and the performance of preventive behaviors could be made. Fourthly, recall and social desirability bias may exist, as only retrospective self-reports of the participants were collected.

Strengths

Despite the above limitations, this is one of few studies to assess the knowledge, perceived beliefs, and preventive behaviors toward COVID-19 among the older adult population in China. Meanwhile, a large number of respondents from all provinces in mainland China were recruited, enabling us to obtain a wide range of participants with various demographic backgrounds. Additionally, the investigation was conducted in the stable epidemic period (February 19 to March 19, 2020), when community-wide COVID-19 prevention activities were launched by local health authorities; thus, we were able to assess how community-level control measures influenced personal preventive behaviors. Therefore, our results are of practical significance for the design and implementation of health programs for the older population in the prevention of COVID-19 and other emerging epidemics.

Conclusions

Generally, the Chinese older adults in our survey demonstrated good knowledge, optimistic perceived beliefs, and appropriate behaviors toward COVID-19 during the pandemic, which are important factors to limit the spread of the disease. Higher behavior scores were found among older adults who were married and cohabitating and who were restricted by community-level control measures, suggesting that health education programs should pay more attention to people who are single and can move around freely. In addition, our findings suggest that good knowledge and appropriate perceived beliefs are associated with high levels of engagement in behaviors to prevent COVID-19. Therefore, to promote preventive behaviors, continued health intervention programs are advised among older people to improve their knowledge in certain aspects, including transmission modes and vaccines against COVID-19; additionally, older people's perception of their own vulnerability to COVID-19 and the effectiveness of COVID-19–related preventive behaviors should be further emphasized.

Acknowledgments

We are grateful to all the survey participants.

Authors' Contributions

HMW and ZC conceived and designed the study. YC and BYC assisted in the questionnaire design and data collection. HC, YL, and HHZ assisted in the data collection. YC and RZ contributed to the statistical analysis and drafted the manuscript; HMW, YC, and RZ finalized the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) Situation Report – 202. World Health Organization. 2020 Aug 09. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200809-covid-19-sitrep-202.pdf?sfvrsn=2c7459f6_2 [accessed 2020-08-09]
2. Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A Novel Coronavirus Emerging in China - Key Questions for Impact Assessment. N Engl J Med 2020 Feb 20;382(8):692-694. [doi: 10.1056/NEJMp2000929] [Medline: 31978293]
3. Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. J Evid Based Med 2020 Feb 12;13(1):3-7 [FREE Full text] [doi: 10.1111/jebm.12376] [Medline: 32048815]
4. Meo SA, Alhowikan AM, Al-Khlaiwi T, Meo IM, Halepoto DM, Iqbal M, et al. Novel coronavirus 2019-nCoV: prevalence, biological and clinical characteristics comparison with SARS-CoV and MERS-CoV. Eur Rev Med Pharmacol Sci 2020 Feb;24(4):2012-2019 [FREE Full text] [doi: 10.26355/eurrev_202002_20379] [Medline: 32141570]
5. 2019-nCo VIAEOIC. 2019-nCoV outbreak is an emergency of international concern. World Health Organization. 2020 Jan 32. URL: http://www.euro.who.int/en/health-topics/emergencies/pages/news/news/2020/01/2019-ncov-outbreak-is-an-emergency-of-international-concern [accessed 2020-01-31]
6. WHO Director-General's opening remarks at the media briefing on COVID-19. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-05-11]
7. Rajagopalan S. Tuberculosis and aging: a global health problem. Clin Infect Dis 2001 Oct 01;33(7):1034-1039. [doi: 10.1086/322671] [Medline: 11528577]
8. Nikolich-Zugich J, Knox KS, Rios CT, Natt B, Bhattacharya D, Fain MJ. SARS-CoV-2 and COVID-19 in older adults: what we may expect regarding pathogenesis, immune responses, and outcomes. Geroscience 2020 Apr 10;42(2):505-514 [FREE Full text] [doi: 10.1007/s11357-020-00186-0] [Medline: 32274617]
9. Banerjee D. The impact of Covid-19 pandemic on elderly mental health. Int J Geriatr Psychiatry 2020 May 04 [FREE Full text] [doi: 10.1002/gps.5320] [Medline: 32364283]
10. Liu K, Chen Y, Lin R, Han K. Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. J Infect 2020 Jun;80(6):e14-e18 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.005] [Medline: 32171866]
11. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 2020 May 12;323(18):1775-1776. [doi: 10.1001/jama.2020.4683] [Medline: 32203977]
12. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
13. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. Article in Chinese. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Feb 10;41(2):145-151. [doi: 10.3760/cma.j.issn.0254-6450.2020.02.003] [Medline: 32064853]
14. Visentin LM, Bondy SJ, Schwartz B, Morrison LJ. Use of personal protective equipment during infectious disease outbreak and nonoutbreak conditions: a survey of emergency medical technicians. CJEM 2009 Jan 21;11(1):44-56. [doi: 10.1017/s1481803500010915] [Medline: 19166639]
15. Seto WH, Tsang D, Yung RWH, Ching TY, Ng TK, Ho M, Advisors of Expert SARS group of Hospital Authority. Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS). Lancet 2003 May 03;361(9368):1519-1520 [FREE Full text] [doi: 10.1016/s0140-6736(03)13168-6] [Medline: 12737864]
16. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020 Mar;55(3):105924 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105924] [Medline: 32081636]
17. Alimohamadi Y, Taghdir M, Sepandi M. Estimate of the Basic Reproduction Number for COVID-19: A Systematic Review and Meta-analysis. J Prev Med Public Health 2020 May;53(3):151-157 [FREE Full text] [doi: 10.3961/jpmph.20.076] [Medline: 32498136]
18. COVID-19 Guidance for Older Adults. US Centers for Disease Control and Prevention. 2020 Sep 11. URL: https://www.cdc.gov/aging/covid19-guidance.html [accessed 2020-09-11]
19. Barber SJ, Kim H. COVID-19 Worries and Behavior Changes in Older and Younger Men and Women. J Gerontol B Psychol Sci Soc Sci 2020 May 19:gbaa068 [FREE Full text] [doi: 10.1093/geronb/gbaa068] [Medline: 32427341]
20. Liao Q, Cowling B, Lam WT, Ng MW, Fielding R. Situational awareness and health protective responses to pandemic influenza A (H1N1) in Hong Kong: a cross-sectional study. PLoS One 2010 Oct 12;5(10):e13350 [FREE Full text] [doi: 10.1371/journal.pone.0013350] [Medline: 20967280]
21. Kamate SK, Agrawal A, Chaudhary H, Singh K, Mishra P, Asawa K. Public knowledge, attitude and behavioural changes in an Indian population during the Influenza A (H1N1) outbreak. J Infect Dev Ctries 2009 Nov 30;4(1):7-14 [FREE Full text] [doi: 10.3855/jidc.501] [Medline: 20130372]

22. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. BMJ 2009 Jul 02;339:b2651 [FREE Full text] [doi: 10.1136/bmj.b2651] [Medline: 19574308]
23. Yıldırım M, Geçer E, Akgül Ö. The impacts of vulnerability, perceived risk, and fear on preventive behaviours against COVID-19. Psychol Health Med 2020 Jun 03:1-9. [doi: 10.1080/13548506.2020.1776891] [Medline: 32490689]
24. Gesser-Edelsburg A, Cohen R, Hijazi R, Abed Elhadi Shahbari N. Analysis of Public Perception of the Israeli Government's Early Emergency Instructions Regarding COVID-19: Online Survey Study. J Med Internet Res 2020 May 15;22(5):e19370 [FREE Full text] [doi: 10.2196/19370] [Medline: 32392172]
25. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
26. Carpenter CJ. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010 Dec;25(8):661-669. [doi: 10.1080/10410236.2010.521906] [Medline: 21153982]
27. Jones CJ, Smith H, Llewellyn C. Evaluating the effectiveness of health belief model interventions in improving adherence: a systematic review. Health Psychol Rev 2014;8(3):253-269. [doi: 10.1080/17437199.2013.802623] [Medline: 25053213]
28. Gu Y, Li L, Zhou C, Yang T, Dong H. Factors influencing voluntary premarital medical examination in Zhejiang province, China: a culturally-tailored health behavioral model analysis. BMC Public Health 2014 Jun 28;14:659 [FREE Full text] [doi: 10.1186/1471-2458-14-659] [Medline: 24972866]
29. Wong C, Tang CS. Practice of habitual and volitional health behaviors to prevent severe acute respiratory syndrome among Chinese adolescents in Hong Kong. J Adolesc Health 2005 Mar;36(3):193-200 [FREE Full text] [doi: 10.1016/j.jadohealth.2004.02.024] [Medline: 15737774]
30. Quinn SC, Kumar S, Freimuth VS, Kidwell K, Musa D. Public willingness to take a vaccine or drug under Emergency Use Authorization during the 2009 H1N1 pandemic. Biosecur Bioterror 2009 Sep;7(3):275-290 [FREE Full text] [doi: 10.1089/bsp.2009.0041] [Medline: 19775200]
31. Tang CSK, Wong C. An outbreak of the severe acute respiratory syndrome: predictors of health behaviors and effect of community prevention measures in Hong Kong, China. Am J Public Health 2003 Nov;93(11):1887-1888. [doi: 10.2105/ajph.93.11.1887] [Medline: 14600058]
32. Lau JTF, Yang X, Tsui H, Pang E, Kim JH. SARS preventive and risk behaviours of Hong Kong air travellers. Epidemiol Infect 2004 Aug;132(4):727-736. [doi: 10.1017/s0950268804002225] [Medline: 15310175]
33. Tang CS, Wong C. Psychosocial factors influencing the practice of preventive behaviors against the severe acute respiratory syndrome among older Chinese in Hong Kong. J Aging Health 2005 Aug;17(4):490-506. [doi: 10.1177/0898264305277966] [Medline: 16020576]
34. Clements JM. Knowledge and Behaviors Toward COVID-19 Among US Residents During the Early Days of the Pandemic: Cross-Sectional Online Questionnaire. JMIR Public Health Surveill 2020 May 08;6(2):e19161 [FREE Full text] [doi: 10.2196/19161] [Medline: 32369759]
35. Niu Z, Wang T, Hu P, Mei J, Tang Z. Chinese Public's Engagement in Preventive and Intervening Health Behaviors During the Early Breakout of COVID-19: Cross-Sectional Study. J Med Internet Res 2020 Aug 21;22(8):e19995 [FREE Full text] [doi: 10.2196/19995] [Medline: 32716897]
36. Bashirian S, Jenabi E, Khazaei S, Barati M, Karimi-Shahanjarini A, Zareian S, et al. Factors associated with preventive behaviours of COVID-19 among hospital staff in Iran in 2020: an application of the Protection Motivation Theory. J Hosp Infect 2020 Jul;105(3):430-433 [FREE Full text] [doi: 10.1016/j.jhin.2020.04.035] [Medline: 32360337]
37. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental Health and Behavior of College Students During the Early Phases of the COVID-19 Pandemic: Longitudinal Smartphone and Ecological Momentary Assessment Study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
38. Zhang M, Li Q, Du X, Zuo D, Ding Y, Tan X, et al. Health Behavior Toward COVID-19: The Role of Demographic Factors, Knowledge, and Attitude Among Chinese College Students During the Quarantine Period. Asia Pac J Public Health 2020 Aug 19:1010539520951408. [doi: 10.1177/1010539520951408] [Medline: 32814434]
39. Guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (The Third Trial Version). General Office of the National Health Commission of the People's Republic of China. 2020. URL: http://www.nhc.gov.cn/yzygj/s7653p/202001/f492c9153ea9437bb587ce2ffcbee1fa.shtml [accessed 2020-01-23]
40. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
41. Clements JM. Knowledge and Behaviors Toward COVID-19 Among US Residents During the Early Days of the Pandemic: Cross-Sectional Online Questionnaire. JMIR Public Health Surveill 2020 May 08;6(2):e19161 [FREE Full text] [doi: 10.2196/19161] [Medline: 32369759]
42. Jehn M, Kim Y, Bradley B, Lant T. Community knowledge, risk perception, and preparedness for the 2009 influenza A/H1N1 pandemic. J Public Health Manag Pract 2011;17(5):431-438. [doi: 10.1097/PHH.0b013e3182113921] [Medline: 21788781]

43. Alanzi M, Albalawi M, Kabrah S, Aljehani Y, Okashah A, Aljohani Z, et al. Knowledge, Attitudes, and Practices (KAPs) of Healthcare Workers towards MERS-CoV Infection at PHCs in Madinah, KSA during Hajj 1440, 2019. Am J Microbiol Res 2019;7(4):122-129 [FREE Full text]
44. Lin Y, Huang L, Nie S, Liu Z, Yu H, Yan W, et al. Knowledge, attitudes and practices (KAP) related to the pandemic (H1N1) 2009 among Chinese general population: a telephone survey. BMC Infect Dis 2011 May 16;11:128 [FREE Full text] [doi: 10.1186/1471-2334-11-128] [Medline: 21575222]
45. Lau JTF, Griffiths S, Choi KC, Tsui HY. Widespread public misconception in the early phase of the H1N1 influenza epidemic. J Infect 2009 Aug;59(2):122-127. [doi: 10.1016/j.jinf.2009.06.004] [Medline: 19592114]
46. Huang Y, Wu Q, Wang P, Xu Y, Wang L, Zhao Y, et al. Measures Undertaken in China to Avoid COVID-19 Infection: Internet-Based, Cross-Sectional Survey Study. J Med Internet Res 2020 May 12;22(5):e18718 [FREE Full text] [doi: 10.2196/18718] [Medline: 32396516]
47. Doogan C, Buntine W, Linger H, Brunt S. Public Perceptions and Attitudes Toward COVID-19 Nonpharmaceutical Interventions Across Six Countries: A Topic Modeling Analysis of Twitter Data. J Med Internet Res 2020 Sep 03;22(9):e21419 [FREE Full text] [doi: 10.2196/21419] [Medline: 32784190]
48. Kong H, Xiao Q, Yang M, Zu W. Awareness on knowledge about coronavirus disease 2019 prevention and control among residents in Chaoyang district of Beijing City. Article in Chinese. Chin J Public Health 2020 Apr 30;36(6):959-962. [doi: 10.11847/zgggws1129479]
49. Fung IC, Cairncross S. Effectiveness of handwashing in preventing SARS: a review. Trop Med Int Health 2006 Nov;11(11):1749-1758 [FREE Full text] [doi: 10.1111/j.1365-3156.2006.01734.x] [Medline: 17054756]
50. De Wandel D, Maes L, Labeau S, Vereecken C, Blot S. Behavioral determinants of hand hygiene compliance in intensive care units. Am J Crit Care 2010 May;19(3):230-239. [doi: 10.4037/ajcc2010892] [Medline: 20436062]
51. Lau JTF, Yang X, Tsui HY, Kim JH. Impacts of SARS on health-seeking behaviors in general population in Hong Kong. Prev Med 2005 Aug;41(2):454-462 [FREE Full text] [doi: 10.1016/j.ypmed.2004.11.023] [Medline: 15917041]
52. Curtis VA, Danquah LO, Aunger RV. Planned, motivated and habitual hygiene behaviour: an eleven country review. Health Educ Res 2009 Aug;24(4):655-673 [FREE Full text] [doi: 10.1093/her/cyp002] [Medline: 19286894]
53. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
54. Jones JH, Salathé M. Early assessment of anxiety and behavioral response to novel swine-origin influenza A(H1N1). PLoS One 2009 Dec 03;4(12):e8032 [FREE Full text] [doi: 10.1371/journal.pone.0008032] [Medline: 19997505]
55. Barr M, Raphael B, Taylor M, Stevens G, Jorm L, Giffin M, et al. Pandemic influenza in Australia: using telephone surveys to measure perceptions of threat and willingness to comply. BMC Infect Dis 2008 Sep 15;8:117 [FREE Full text] [doi: 10.1186/1471-2334-8-117] [Medline: 18793441]
56. Eastwood K, Durrheim D, Francis JL, d'Espaignet ET, Duncan S, Islam F, et al. Knowledge about pandemic influenza and compliance with containment measures among Australians. Bull World Health Organ 2009 Aug;87(8):588-594 [FREE Full text] [doi: 10.2471/blt.08.060772] [Medline: 19705008]
57. Tang CS, Wong C. Factors influencing the wearing of facemasks to prevent the severe acute respiratory syndrome among adult Chinese in Hong Kong. Prev Med 2004 Dec;39(6):1187-1193 [FREE Full text] [doi: 10.1016/j.ypmed.2004.04.032] [Medline: 15539054]
58. Lau JTF, Kim JH, Tsui HY, Griffiths S. Anticipated and current preventive behaviors in response to an anticipated human-to-human H5N1 epidemic in the Hong Kong Chinese general population. BMC Infect Dis 2007 Mar 15;7:18 [FREE Full text] [doi: 10.1186/1471-2334-7-18] [Medline: 17359545]
59. Hsu HY, Gallinagh R. The relationships between health beliefs and utilization of free health examinations in older people living in a community setting in Taiwan. J Adv Nurs 2001 Sep;35(6):864-873. [doi: 10.1046/j.1365-2648.2001.01924.x] [Medline: 11555034]
60. Lagerlund M, Hedin A, Sparén P, Thurfjell E, Lambe M. Attitudes, beliefs, and knowledge as predictors of nonattendance in a Swedish population-based mammography screening program. Prev Med 2000 Oct;31(4):417-428. [doi: 10.1006/pmed.2000.0723] [Medline: 11006068]
61. Nie S, Cao J, Tuo A, Zheng Y. Public's status of KAP for COVID-19 and its influencing factors. Article in Chinese. Shanghai Journal of Preventive Medicine 2020 Jul;7(6). [doi: 10.19428/j.cnki.sjpm.2020.20218]
62. Bronfenbrenner U. Toward an experimental ecology of human development. Am Psychol 1977;32(7):513-531. [doi: 10.1037/0003-066x.32.7.513]
63. Wenzel RP, Edmond MB. Managing SARS amidst uncertainty. N Engl J Med 2003 May 15;348(20):1947-1948. [doi: 10.1056/NEJMp030072] [Medline: 12748313]
64. Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. Soc Sci Med 2000 Sep;51(6):843-857. [doi: 10.1016/s0277-9536(00)00065-4] [Medline: 10972429]
65. Leung GM, Quah S, Ho L, Ho S, Hedley AJ, Lee H, et al. A tale of two cities: community psychobehavioral surveillance and related impact on outbreak control in Hong Kong and Singapore during the severe acute respiratory syndrome epidemic. Infect Control Hosp Epidemiol 2004 Dec;25(12):1033-1041. [doi: 10.1086/502340] [Medline: 15636289]

66. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
67. Lau JTF, Griffiths S, Au DWH, Choi KC. Changes in knowledge, perceptions, preventive behaviours and psychological responses in the pre-community outbreak phase of the H1N1 epidemic. Epidemiol Infect 2011 Jan;139(1):80-90. [doi: 10.1017/S0950268810001925] [Medline: 20800008]
68. Wise T, Zbozinek T, Michelini G, Hagan C. Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States. PsyArXiv Preprint posted online on March 19, 2020. [doi: 10.31234/osf.io/dz428]
69. Jernigan JC, Trauth JM, Neal-Ferguson D, Cartier-Ulrich C. Factors that influence cancer screening in older African American men and women: focus group findings. Fam Community Health 2001 Oct;24(3):27-33. [doi: 10.1097/00003727-200110000-00005] [Medline: 11563942]
70. Wardle J, Williamson S, McCaffery K, Sutton S, Taylor T, Edwards R, et al. Increasing attendance at colorectal cancer screening: testing the efficacy of a mailed, psychoeducational intervention in a community sample of older adults. Health Psychol 2003 Jan;22(1):99-105. [doi: 10.1037//0278-6133.22.1.99] [Medline: 12558207]
71. Booth M, Bauman A, Owen N. Perceived barriers to physical activity among older Australians. J Aging Phys Act 2020 Jul 01;10(3):271-280. [doi: 10.1123/japa.10.3.271]
72. Lau JTF, Yang X, Tsui HY, Pang E. SARS related preventive and risk behaviours practised by Hong Kong-mainland China cross border travellers during the outbreak of the SARS epidemic in Hong Kong. J Epidemiol Community Health 2004 Dec;58(12):988-996 [FREE Full text] [doi: 10.1136/jech.2003.017483] [Medline: 15547057]
73. Lau JTF, Yang X, Tsui H, Kim JH. Monitoring community responses to the SARS epidemic in Hong Kong: from day 10 to day 62. J Epidemiol Community Health 2003 Nov;57(11):864-870 [FREE Full text] [doi: 10.1136/jech.57.11.864] [Medline: 14600111]
74. Smith Klohn L, Rogers RW. Dimensions of the severity of a health threat: the persuasive effects of visibility, time of onset, and rate of onset on young women's intentions to prevent osteoporosis. Health Psychol 1991;10(5):323-329. [doi: 10.1037//0278-6133.10.5.323] [Medline: 1935867]
75. Aoun S, Donovan RJ, Johnson L, Egger G. Preventive Care in the Context of Men's Health. J Health Psychol 2002 May;7(3):243-252. [doi: 10.1177/1359105302007003219] [Medline: 22114248]
76. Ronis DL. Conditional health threats: health beliefs, decisions, and behaviors among adults. Health Psychol 1992;11(2):127-134. [doi: 10.1037//0278-6133.11.2.127] [Medline: 1582381]

Abbreviations

ANOVA: analysis of variance
CDC: Center for Disease Control and Prevention
HBM: health belief model
MERS: Middle East respiratory syndrome
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by G Eysenbach; submitted 21.08.20; peer-reviewed by T Yang, A Al-Hasan; comments to author 26.08.20; revised version received 20.10.20; accepted 07.12.20; published 31.12.20.

Please cite as:
Chen Y, Zhou R, Chen B, Chen H, Li Y, Chen Z, Zhu H, Wang H
Knowledge, Perceived Beliefs, and Preventive Behaviors Related to COVID-19 Among Chinese Older Adults: Cross-Sectional Web-Based Survey
J Med Internet Res 2020;22(12):e23729
URL: https://www.jmir.org/2020/12/e23729
doi:10.2196/23729
PMID:33293262

Original Paper

Association of Social Media Use With Mental Health Conditions of Nonpatients During the COVID-19 Outbreak: Insights from a National Survey Study

Bu Zhong[1,2], DPhil; Zhibin Jiang[3], DPhil; Wenjing Xie[4], DPhil; Xuebing Qin[5], DPhil

[1]Donald P Bellisario College of Communications, Pennsylvania State University, University Park, PA, United States
[2]School of Journalism and Communication, South China University of Technology, Guangzhou, China
[3]School of Journalism and Communication, Shanghai International Studies University, Shanghai, China
[4]School of Communication and the Arts, Marist College, Poughkeepsie, NY, United States
[5]School of Communication, East China Normal University, Shanghai, China

Corresponding Author:
Xuebing Qin, DPhil
School of Communication
East China Normal University
500 Dongchuan Road
Minhang District
Shanghai, 200062
China
Phone: 86 13818950148
Email: xbqin@comm.ecnu.edu.cn

Abstract

Background: Considerable research has been devoted to examining the mental health conditions of patients with COVID-19 and medical staff attending to these patients during the COVID-19 pandemic. However, there are few insights concerning how the pandemic may take a toll on the mental health of the general population, and especially of nonpatients (ie, individuals who have not contracted COVID-19).

Objective: This study aimed to investigate the association between social media use and mental health conditions in the general population based on a national representative sample during the peak of the COVID-19 outbreak in China.

Methods: We formed a national representative sample (N=2185) comprising participants from 30 provinces across China, who were the first to experience the COVID-19 outbreak in the world. We administered a web-based survey to these participants to analyze social media use, health information support received via social media, and possible psychiatric disorders, including secondary traumatic stress (STS) and vicarious trauma (VT).

Results: Social media use did not cause mental health issues, but it mediated the levels of traumatic emotions among nonpatients. Participants received health information support via social media, but excessive social media use led to elevated levels of stress (β=.175; $P<.001$), anxiety (β=.224; $P<.001$), depression (β=.201; $P<.001$), STS (β=.307; $P<.001$), and VT (β=.688; $P<.001$). Geographic location (or geolocation) and lockdown conditions also contributed to more instances of traumatic disorders. Participants living in big cities were more stressed than those living in rural areas (P=.02). Furthermore, participants from small cities or towns were more anxious (P=.01), stressed ($P<.001$), and depressed (P=.008) than those from rural areas. Obtaining more informational support (β=.165; $P<.001$) and emotional support (β=.144; $P<.001$) via social media increased their VT levels. Peer support received via social media increased both VT (β=.332; $P<.001$) and STS (β=.130; $P<.001$) levels. Moreover, geolocation moderated the relationships between emotional support on social media and VT (F_2=3.549; P=.029) and the association between peer support and STS (F_2=5.059; P=.006). Geolocation also interacted with health information support in predicting STS (F_2=5.093; P=.006).

Conclusions: COVID-19 has taken a severe toll on the mental health of the general population, including individuals who have no history of psychiatric disorders or coronavirus infection. This study contributes to the literature by establishing the association between social media use and psychiatric disorders among the general public during the COVID-19 outbreak. The study findings suggest that the causes of such psychiatric disorders are complex and multifactorial, and social media use is a potential factor.

The findings also highlight the experiences of people in China and can help global citizens and health policymakers to mitigate the effects of psychiatric disorders during this and other public health crises, which should be regarded as a key component of a global pandemic response.

(J Med Internet Res 2020;22(12):e23696) doi:10.2196/23696

KEYWORDS

COVID-19; mental health; social media; health information support; secondary traumatic stress; vicarious trauma; social support

Introduction

Background

After COVID-19 hit the world, health care workers have been rushing to care for infected patients and save lives. In the race to contain the COVID-19 pandemic, it is important for health care providers to not ignore another big risk—the invisible toll on mental health among nonpatients. One of the key lessons we have learned from the COVID-19 pandemic is that living in this public health crisis is extremely stressful for everyone, including those who appear healthy and have not contracted the virus [1]. This was evident in China, where fear, worry, and anxiety about COVID-19 and its health risks were overwhelming as soon as the disease spread across the country [2]. Following the lockdown announced in Wuhan on January 23, 2020, other Chinese cities quickly restricted people from moving around. A mandatory lockdown had forced millions of Chinese to stay at home for weeks, and even months. Living in an isolated environment could make people feel unattached, worried, lonely, and even traumatized [3]. The strong emotions experienced by the Chinese during the initial months of the COVID-19 outbreak were almost unimaginable to the outside world until it spiraled into a global pandemic [2], resulting in millions of cases and hundreds and thousands of deaths worldwide.

As researchers in epidemiology, medicine, and public health worldwide are continuously researching medications, vaccines, and coping strategies for COVID-19, it is important to study how health information on social media and lockdown or quarantine situations may contribute to the toll on people's mental health. Many studies addressing the impact of the pandemic on mental health focus on COVID-19 patients [4], who have shown, for instance, posttraumatic stress symptoms [5] or depression [6]. This study aimed to reveal a holistic picture by evaluating the effects of the COVID-19 pandemic on the mental health of nonpatients based on a national representative sample from China. Using a national representative sample is critical for such a study given the widespread urban-rural differences in China. The COVID-19 pandemic has had different effects across China's cities, towns, and rural areas, as the local responses in these regions depended on a complex interplay of numerous social and economic factors. Hence, the toll on mental health on people in these regions could considerably vary depending on their geographic location (or geolocation) and other internal or external factors.

Social Media Use and COVID-19

Considerable research has been devoted to examining the mental health of patients diagnosed with COVID-19 [5-7] and medical staff who cared for and treated these patients during the pandemic [8,9]. However, not much is known about the mental health conditions of nonpatients, who might experience varying degrees of psychiatric disorders due to COVID-19. Research shows that as people are consistently exposed to negative information about a crisis, their anxiety and depression levels could elevate for an extended period [10]. As COVID-19 spread in China, people began to use WeChat, China's leading social media app, more frequently. By February 8, 2020, over 100 "mini programs" were added to WeChat to provide epidemic status information, and the app's users grew by nearly 60% within 3 weeks [11]. The growing use of social media for crisis management has been well documented in the literature, and social media is considered a powerful tool to share health information related to pandemic risks [12,13].

However, there are controversies concerning the links between social media use and mental health. Studies have found that social media use may decrease satisfaction of life [14] and increase self-harm, suicidal ideation [15,16], psychological distress, depression, and anxiety [14]. Systematic reviews have shown that most of these studies are from Western countries, and a few studies are from Asian countries [14,15]. Insights from Asian countries may not only deepen the understanding of the relationship between social media use and mental health but also provide suggestions for education and policies [14]. People benefit from using social media in terms of promoting behavior change [17], obtaining health information support, and staying connected with others [2]; however, social media could also spread fear or misinformation about COVID-19, thereby causing harm to their mental health and psychological well-being [13]. Thus, more effort is merited to study how using social media to seek and share health information could have an impact on the users' mental health during a health crisis.

Informational, Emotional, and Peer Support

One of the main reasons people have been sharing health information on social media during the COVID-19 outbreak is the social support they gain from other users—a phenomenon that can be best explained by the uses and gratifications theory [18,19]. This theory holds that people use certain media content or platforms to gratify specific informational needs and demands; otherwise, they would no longer come back and use it again. Health information provides significant social support to people with health concerns, resulting in a number of benefits that help symptom control, disease recovery, life safety, and overall well-being [20]. Social support is defined as "the individual feeling valued and cared for by their social network as well as how well the person is embedded into a network of communication and social obligation" [21]. In other words, social support refers to the perception that one is cared for and support is exchanged through interpersonal interactions [22].

Helgeson [23] argues that social support has 3 main forms: informational, emotional, and instrumental support. Informational support refers to the provision of advice, guidance, and other useful information [24]. Emotional support exhibits the expressions of care, concern, empathy, and sympathy [23]. Instrumental support represents the concrete and direct ways in which people assist others [25,26]; this is referred to as "peer support" in this study.

The support users obtain on social media through accessing and sharing pandemic-related information can be viewed as health information support [2]. It functions like a type of social support received from family members, friends, colleagues, or peers during the pandemic. Ample research has identified that health information support provides patients with significant care and emotional support [27]. This type of support can also improve users' capability of making informed medical decisions [28]. Identifying with social media groups has been found to increase one's self-esteem and self-efficacy, and thus reduce uncertainty about the self [29]. Seeking support and social connection is a critical point in the lives of people with chronic conditions [30].

In addition to informational support, social media users gain emotional and peer support from the health information they access [2]. Emotional support is a key component of peer support in health care settings. High emotional support is known to mitigate the stress response and prevent consequent adverse effects on the progression of depression, posttraumatic stress disorder, and prostate cancer [31]. Patients who report more tangible emotional support are more likely to have experienced a positive social interaction with fellow patients and medical professionals [32]. WeChat users can obtain emotional support by chatting with friends, joining social media groups of their interests, and staying connected with others [20].

Peer support is a subcategory of social support, and it is differentiated by the source of support received from peers who are in a similar demographic group or illness community. Social media groups are important platforms to discuss medical conditions, share personal experiences, and seek health information [33,34]. In this study, peer support is defined as a type of support social media users receive from others when they share their knowledge and experiences of COVID-19. However, the quality of health information on the internet may vary depending on the sources, and health misinformation has become a severe threat to public health [35]. For instance, the inaccuracy of online health information deteriorates the physician-patient relationship and erodes trust in doctors [20]. Thus, it would be important to study how the social support about COVID-19 that people receive on social media is associated with their mental health and psychiatric disorders. Overall, peer support complements and enhances mental health by providing the necessary emotional, social, and practical assistance for managing disease and staying healthy.

Mental Health and the Pandemic

Global health crises, especially the COVID-19 pandemic, have diverse and substantial health implications on the human society [1]. The deleterious consequences related to COVID-19 along with the unprecedented mitigation strategies pose major threats to the well-being and mental health of people worldwide [36]. COVID-19 has significantly changed various aspects of our routine life, including economics, travel, interpersonal communication, and health management [37]. When an individual's routine life is suddenly and severely disrupted by a pandemic, the human brain may no longer function normally as usual; this can, consequently, lead to stress or psychiatric disorders [38]. This affects not only the people who have been diagnosed with COVID-19 but also those who seem to be "normal" or "healthy." Thus, a psychiatric disorder may occur in individuals who have not been infected by the virus themselves but have experienced or witnessed the challenges of other individuals during the COVID-19 pandemic [39]. Although findings regarding the relationship between social support and mental health are inconsistent, social support, in general, has been found to provide physical and psychological advantages to combat stressful events and recover from psychological distress [40]. Lack of social support, however, has been linked with the onset and development of depression [41], mood disorders [42], and other medical illnesses such as multiple sclerosis and rheumatoid arthritis [43,44].

Research Questions

Building upon previous findings, this study is among the first to investigate nonpatients' mental health conditions such as stress, anxiety, depression, secondary traumatic stress (STS), and vicarious trauma (VT) during the peak of the COVID-19 epidemic in China. We measured both STS and VT, but not posttraumatic stress symptoms, as COVID-19 was an ongoing crisis when we collected data for analyses. Thereafter, we analyzed the internal (ie, demographics) and external (ie, pandemic and environmental conditions) factors that may contribute to possible psychiatric disorders such as STS and VT. Specifically, we investigated the following research questions (RQs):

- RQ1: Did nonpatients experience any stress, anxiety, depression, STS, or VT at the peak of the COVID-19 outbreak in China?
- RQ2: Did people living in different geolocations experience varying levels of psychiatric disorders?
- RQ3: How did demographics and pandemic situations, such as lockdown, quarantine conditions, and death numbers, contribute to possible psychiatric disorders?
- RQ4: How was social media use associated with psychiatric disorders?
- RQ5: How did health informational, emotional, and peer support mediate the relationship between demographics or pandemic situations and STS or VT?
- RQ6: Would geolocation interact with the health informational, emotional, and peer support people received through social media to predict psychiatric disorders?

Methods

Sampling

A marketing research company helped recruit a national representative sample for this study by using the quota sampling method, with a survey investigating how social media use affected the mental health conditions of people in China. In all, 4500 questionnaires were distributed, and 3820 individuals participated in this survey. After excluding incomplete data, responses from a total of 2185 participants were included for further analyses. These participants were from 30 provinces across China, and lived in big cities (eg, Beijing, Shanghai, Guangzhou, and Wuhan), small cities or towns, and rural areas. To avoid retraumatizing participants with questions about mental health issues or COVID-19, the survey comprised screening questions to exclude patients with COVID-19 or those who currently or previously had depressive or traumatic disorders.

After obtaining approval from the university's institutional review board, we conducted a web-based survey in February 2020, which was the peak of the COVID-19 epidemic in China when most people were forced to live under lockdown conditions due to the spike of confirmed cases. For this survey, participants received a small financial incentive (¥10 or US $1.34 each) and were asked to answer specific questions related to demographics, social media use, changes in mental health conditions, and lockdown conditions. Participants who indicated interest in this study were sent a message that contained the survey URL and login credentials. The survey was password-protected and could not be accessed without these credentials.

Measures

Stress, anxiety, and depression were measured using a 4-point Likert scale, with scores ranging from 0 ("Did not apply to me at all") to 3 ("Applied to me very much, or most of the time"); all other measures were assessed using a 5-point Likert scale, with scores ranging from 1 ("strongly disagree") to 5 ("strongly agree").

Social Media Usage

WeChat is China's dominant social media app, and participants' use of this app was measured using a 6-item instrument that was originally developed for measuring Facebook addiction [45]. The wordings of the questionnaire items were slightly revised to better fit the participants' actual WeChat use during the COVID-19 outbreak. For instance, participants were asked to rate how likely they agreed with 6 statements, including "You feel an urge to use WeChat more as you want to know more about the epidemic" and "You use WeChat for health information on the epidemic so much that it has had a negative impact on your life." The scores on the 6 items were averaged to form the index of social media use (mean 3.123, SD 0.809; Cronbach α=.901). A higher value indicates excessive social media use.

Informational, Emotional, and Peer Support

The scales of informational, emotional, and peer support were adopted and revised based on previous studies [46,47]. Informational support was measured on the basis of 4 items, including "If I have a question or need help related to the coronavirus epidemic, I can usually find the answers on WeChat." The scores on these 4 items were averaged to form the informational support index (mean 3.376, SD 0.900; Cronbach α=.868). Emotional support was measured on the basis of 4 items, including "The health information on WeChat helps me alleviate feelings of loneliness." The emotional support index had a high level of internal consistency (mean 3.292, SD 0.892; Cronbach α=.908). Similarly, peer support was measured on the basis of 6 items, including "WeChat friends give me additional information about the coronavirus epidemic that I am not familiar with." The peer support index also had a high internal consistency (mean 3.245, SD 0.586; Cronbach α=.907).

Stress, Anxiety, and Depression

We used 7 items of the 21-item Depression Anxiety Stress Scale (DASS-21) [48] to measure stress, including "I felt that I was using a lot of nervous energy" and "I found myself getting agitated." These questionnaire items were evaluated on a 4-point Likert scale, with scores ranging from 0 ("Did not apply to me at all") to 3 ("Applied to me very much, or most of the time"). The sum of the scores on these 7 items formed the stress index (mean 4.968, SD 4.455; Cronbach α=.860). DASS-21 was used to measure both anxiety (mean 5.030, SD 4.799; Cronbach α=.860) and depression (mean 5.104, SD 4.975; Cronbach α=.860).

STS and VT

STS is the emotional distress a person experiences when they hear about first-hand trauma experiences of another person [49]. In this study, STS was measured using the 14-item instrument adopted from Bride et al [50], including "I felt emotionally numb" and "My heart started pounding when I thought about the coronavirus epidemic." The scores on these 14 items were averaged to form the STS index (mean 2.466, SD 0.799; Cronbach α=.938). A higher value indicates higher level of STS.

The concept of VT was proposed by Pearlman and Saakvitne [51] in their description of the trauma experiences people have after being exposed to others' trauma stories and having witnessed the pain, fear, and terror that traumatized survivors have endured. In this study, VT was measured using the 8-item instrument developed by Vrklevski and Franklin [52], including "I find myself thinking about distressing material at home" and "Sometimes I feel helpless to assist others in the way I would like." The VT index had a good internal consistency (mean 3.349, SD 0.723; Cronbach α=.861).

Finally, pandemic situations were measured using the following questions: "How long is your residence area locked down to restrict entry of nonresidents?" "Do you have any family members or friends currently under quarantine?" "Do you know any family members or friends confirmed infected by coronavirus?" and "Do you know family members or friends who died due to the coronavirus epidemic?" Data on the participants' age, gender, education, income, and geolocations were also obtained.

Data Analyses

To answer RQ1, we first performed descriptive data analyses with anxiety, depression, stress, STS, and VT. Paired-sample *t* tests were used to compare the levels of different psychiatric disorders. To answer RQ2, we performed one-way analysis of variance with psychiatric disorders as the dependent variables and geolocation as the independent variable. For RQ3 and RQ4, we performed hierarchical regression analyses. STS was entered as the dependent variable in the model, followed by VT, stress, anxiety, and depression. Demographic information was entered in the first step of the model, and pandemic situations were entered in the second step, followed by social media use in the third step. For RQ5, we performed structural equation modeling. Demographics and pandemic situations were used as exogenous variables to predict informational support, emotional support, and peer support, which in turn predicted endogenous variables (ie, STS and VT). For RQ6, we used a generalized linear model to analyze the interaction between geolocation and informational, emotional, and peer support for predicting psychiatric disorders.

Results

Prevalence of Psychiatric Disorders

Compared with the latest census data published by the central government of China [53], the national sample attributes largely matched with the proportion of the Chinese population. The sample demographics were considered as internal factors that may influence the levels of stress, anxiety, or depression (Table 1). Pandemic situations and environmental conditions were considered as external factors that may contribute to STS or VT (Table 2).

The data showed that the pandemic had caused significant harm on people’s mental health. Only 3 weeks after a lockdown was announced in Wuhan, 10% (219/2185) of the participants reported they experienced moderate-to-severe anxiety, and 9.8% (215/2185) of the participants reported they experienced mild anxiety symptoms. Meanwhile, 5.5% (121/2185) of the national sample had moderate-to-severe depression, and 14.5% (316/2185) of the participants reported mild depression. These results are consistent with the findings from other surveys conducted in China during COVID-19 that report approximately 22% of the population experienced anxiety and 20% experienced a combination of depression and anxiety [3], although no study has evaluated traumatic disorders. This study shows that Chinese people displayed a moderate level of STS (mean 2.466, SD 0.799) and a relatively high level of VT (mean 3.934, SD 0.723), with significantly higher VT levels reported than STS levels (t_{2184}=46.747; P<.001).

Table 1. Internal factors contributing to psychiatric disorders (N=2185).

Internal factor	Value
Age (years), mean (SD)	33.43 (31)
Gender, n (%)	2185 (100)
Female	1192 (54.6)
Male	993 (45.6)
Income, n (%)	2185 (100)
Very low income	353 (16.2)
Low income	445 (20.4)
Medium income	1223 (56)
High income	130 (5.9)
Very high income	34 (1.6)
Marital status, n (%)	2185 (100)
Unmarried	715 (32.7)
Married	1470 (67.3)
Education, n (%)	2185 (100)
Middle school or lower	246 (11.3)
High school	492 (22.5)
3-year college	587 (26.9)
4-year college	766 (35.1)
Graduate degree	94 (4.3)

Table 2. External factors contributing to psychiatric disorders (N=2185).

External factor	Value, n (%)
Geolocation	2185 (100)
Rural area	592 (27.1)
Small cities or towns	1189 (54.4)
Big cities	404 (18.5)
Lockdown time	2185 (100)
No lockdown	0 (0)
1-2 weeks	1065 (48.7)
3-4 weeks	795 (36.4)
5-6 weeks	213 (9.7)
7 weeks or more	112 (5.1)
Known quarantine cases	2185 (100)
None	1983 (90.8)
1 person	63 (2.9)
2 people	73 (3.3)
3 people	41 (1.9)
4 or more people	25 (1.1)
Known death cases	2185 (100)
None	2082 (95.3)
1 person	31 (1.4)
2 people	48 (2.2)
3 people	18 (0.8)
4 or more people	6 (0.3)
Known infection cases	2185 (100)
None	2016 (92.3)
1 person	73 (3.3)
2 people	65 (3)
3 people	24 (1.1)
4 or more people	7 (0.3)

Internal and External Factors Contributing to Psychiatric Disorders

Our analyses revealed that a range of internal (ie, demographics) and external (ie, pandemic and environmental conditions) factors are related to psychiatric disorders (Table 3), as well as stress, anxiety, and depression (Table 4). With regard to stress, participants who were younger (β=−.142; *P*<.001), male (β=.054; *P*=.04), married or divorced (β=.078; *P*<.001), and had a higher income (β=.049; *P*=.03) reported higher levels of stress than other participants during the pandemic. These demographics accounted for 2.3% of the variance in stress (ΔR^2=.023; *P*<.001). Higher levels of stress were reported by participants who lived under lockdown for a longer time (β=.028; *P*=.04) and those who knew of more quarantine cases among their friends and family members (β=.105; *P*<.001), as well as more cases of COVID-19–related deaths (β=.117; *P*<.001). These pandemic situations accounted for 4.2% of the variance in stress (ΔR^2=.042; *P*<.001).

With regard to anxiety, participants who were younger (β=−.085; *P*<.001), male (β=.058; *P*=.007), and married or divorced (β=.054; *P*=.03) reported higher levels of anxiety than did the other participants. These demographics accounted for 0.8% of the variance in anxiety (ΔR^2=.008; *P*=.01). Participants who lived longer in a lockdown situation (β=.051; *P*=.02) and those who knew of more quarantine cases among family members and close friends (β=.092; *P*<.001) as well as more cases of deaths among them (β=.085; *P*=.001) reported higher levels of anxiety. These lockdown situations accounted for 3.9% of the variance in anxiety (ΔR^2=.039; *P*<.001).

As for depression, younger participants reported being more depressive than older participants (β=−.094; *P*<.001). Participants who knew of more quarantine cases in family members and close friends (β=.107; *P*<.001) as well as more cases of deaths among them (β=.073; *P*=.02) reported higher levels of depression. These pandemic situations accounted for 3.7% of the variance in depression (ΔR^2=.037; *P*<.001).

With regard to psychiatric disorders such as STS and VT, participants who were younger (β=−.099; *P*<.001), more educated (β=.093; *P*<.001), and married (β=.081; *P*<.001) were more likely to show STS symptoms. These demographics accounted for 1.7% of the variance in STS (ΔR^2=.017; *P*<.001). As these participants knew of more quarantine cases among family members and close friends, they were more likely to develop STS (β=.053; *P*=.048). These pandemic situations accounted for 1.5% of the variance in STS (ΔR^2=.015; *P*<.001). Moreover, female (β=−.059; *P*=.007) and more educated (β=.085; *P*<.001) participants experienced higher levels of VT. These demographics accounted for 1.8% of the variance in VT (ΔR^2=.018; *P*<.001). Knowledge of quarantine cases among family members and close friends also positively predicted VT (β=.057; *P*=.003). These pandemic situations accounted for 0.6% of the variance in VT (ΔR^2=.006; *P*=.01).

Table 3. Factors contributing to secondary traumatic stress and vicarious trauma.

Variables	Secondary traumatic stress				Vicarious trauma			
	β	*t* test (*df*)[a]	*P* value	ΔR^2	β	*t* test (*df*)[a]	*P* value	ΔR^2
Step 1: Demographics				.017[b]				.018[b]
Age	−.099	*−3.899 (5)*[c]	<.001	—[d]	−.013	−.502 (5)	.53	—
Gender	.015	.673 (5)	.54	—	−.059	*−2.705 (5)*	.007	—
Income	−.032	−1.365 (5)	.46	—	.045	1.918 (5)	.05	—
Marriage	.081	*3.197 (5)*	<.001	—	.046	1.810 (5)	.06	—
Education	.093	*3.887 (5)*	<.001	—	.085	*3.570 (5)*	<.001	—
Step 2: Pandemic situations				.015[b]				.006[e]
Lockdown time	.041	1.865 (4)	.07	—	.015	.666 (4)	.54	—
Known quarantine case	.053	*1.976 (4)*	.048	—	.057	*2.122 (4)*	.003	—
Known infected cases	.022	.732 (4)	.46	—	−.026	−.839 (4)	.40	—
Known death cases	.052	1.783 (4)	.095	—	−.006	−.188 (4)	.79	—
Step 3: Social media use				.091[b]				.479[b]
Social media use	.307	*14.899 (1)*	<.001	—	.688	*43.315 (1)*	<.001	—

[a] Two-tailed *t* tests were performed.

[b] This denotes *P*<.001.

[c] Italiziced values indicate statistical significance.

[d] Not applicable.

[e] This denotes *P*<.005.

Table 4. Factors contributing to stress, anxiety, and depression.

Variables	Stress				Anxiety				Depression			
	β	*t* test (*df*)[a]	*P* value	ΔR^2	β	*t* test (*df*)[a]	*P* value	ΔR^2	β	*t* test (*df*)[a]	*P* value	ΔR^2
Step 1 Demographics				.023[b]				.008[c]				.007[c]
Age	−.142	*−5.452 (5)*[d]	<.001	—[e]	−.085	*−3.316 (5)*	<.001	—	−.094	*−3.678 (5)*	<.001	—
Gender	.054	*2.445 (5)*	.04	—	.058	*2.650 (5)*	.007	—	.031	1.419 (5)	.147	—
Income	.049	*2.022 (5)*	.03	—	.017	.729 (5)	.459	—	.019	.782 (5)	.423	—
Married	.078	3.035 (5)	*<.001*	—	.054	*2.412 (5)*	.03	—	.049	1.923 (5)	.125	—
Education	.035	1.424 (5)	.14	—	.005	.210 (5)	.806	—	−.003	−.120 (5)	.930	—
Step 2 Pandemic situations				.042[b]				.039[b]				.037[b]
Lockdown time	.028	*2.438 (4)*	.04	—	.051	*2.329 (4)*	.02	—	.035	1.609 (4)	.114	—
Known quarantine cases	.105	*3.070 (4)*	<.001	—	.092	*3.478 (4)*	<.001	—	.107	*4.045 (4)*	<.001	—
Known infected cases	.006	1.256 (4)	.623	—	.037	1.217 (4)	.223	—	.036	1.168 (4)	.242	—
Known death cases	.117	*3.887 (4)*	<.001	—	.085	*2.901 (4)*	.001	—	.073	*2.483 (4)*	.02	—
Step 3 Social media use				.030[b]				.048[b]				.039[b]
Use of WeChat	.175	*8.175 (1)*	<.001	—	.224	*10.672 (1)*	<.001	—	.201	*9.521 (1)*	<.001	—

[a] Two-tailed *t* tests were performed.

[b] This denotes *P*<.001.

[c] This denotes *P*<.01.

[d] Italiziced values indicate statistical significance.

[e] Not applicable.

Geolocation and Psychiatric Disorders

Participants living in big cities, small cities or towns, and rural areas reported varying levels of stress ($F_{2,2075}$=7.224; *P*<.001). After COVID-19 hit China, participants living in big cities (mean 5.036, SD 4.518) were more stressed than those living in rural areas (mean 4.367, SD 4.351; *P*=.02). Participants living in small cities or towns (mean 5.238, SD 4.460) were also more stressed than those living in rural areas (*P*<.001), but the difference was small between participants living in big cities and those living small cities or towns. A similar pattern was also observed for anxiety, as participants living in these three geolocations reported varying levels of anxiety ($F_{2,2183}$=3.569; *P*=.03). Participants from small cities or towns (mean 5.270, SD 4.747) were more anxious than those from rural areas (mean 4.647, SD 4.820; *P*=.01), but no difference was found between participants from big cities (mean 4.883, SD 4.891) and those from small cities or towns (*P*=.16). Moreover, no significant difference was found between participants from big cities and those from rural areas (*P*=.45).

A similar pattern was observed for depression, concerning the impact of geolocations ($F_{2,2183}$=3.569; *P*=.03). Participants from small cities or towns (mean 5.344, SD 4.877) experienced more depression than those from rural areas (mean 4.682, SD 5.092; *P*=.008). No significant difference was observed between participants living in small and big cities (mean 5.015, SD 5.054; *P*=.25) nor between participants living in big cities and rural areas (*P*=.30).

Social Media Use and Psychiatric Disorders

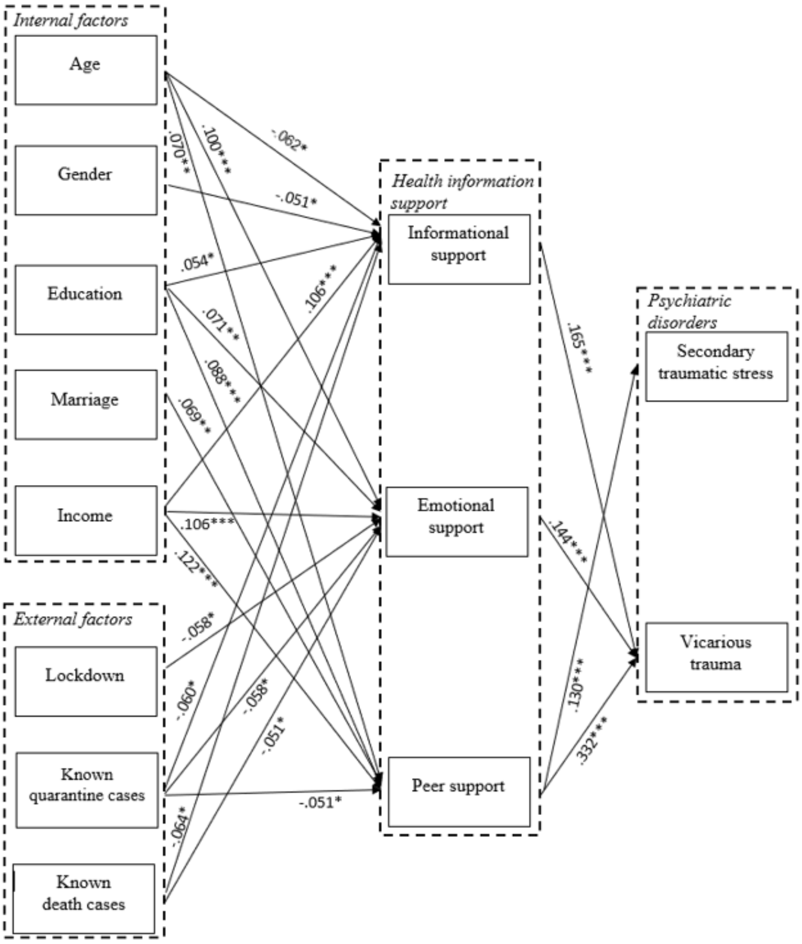
More social media use contributed to STS (β=.307; *P*<.001; ΔR^2=.091; sum of R^2=.124) and VT (β=.688; *P*<.001; ΔR^2=.479; sum of R^2=.481), as shown in Table 3. Participants who used more social media also reported higher levels of stress (β=.175; *P*<.001; ΔR^2=.030; sum of R^2=.095), anxiety (β=.224; *P*<.001; ΔR^2=048; sum of R^2=.095), and depression (β=.201; *P*<.001; ΔR^2=.039; sum of R^2=.083).

Mediating Effects of Informational, Emotional, and Peer Support

Finally, structural equation modeling was performed to evaluate the mediating effects of informational, emotional, and peer support that the participants gained from health information shared on social media. The model had a good fit as demonstrated by the following indices: χ^2_{19}=25.286, minimum discrepancy divided by its *df* (CMIN/*df*)=1.331, *P*=.15, root mean square error of approximation=0.012, comparative fit index=.999, and Bentler-Bonett Normed fit index=0.996. Figure 1 shows that participants who were younger (β=−.062; *P*=.03), female (β=−.051; *P*=.05), and more educated (β=.054; *P*=.05) and those who earned a higher income (β=.106; *P*<.001) received more informational support on social media. On the other hand, participants who were older (β=.100; *P*<.001) and more educated (β=.071; *P*=.007) and those who had a higher income (β=.106; *P*<.001) received more emotional support in using social media. Participants who were older (β=.070; *P*=.005), married (β=.069; *P*=.004), and more educated (β=.088; *P*<.001) and those who had a higher income (β=.122; *P*<.001) received more peer support in using social media.

Knowledge of more quarantine cases among family members and close friends could negatively affect informational support (β=−.060; *P*=.04), emotional support (β=−.058; *P*=.03), or peer support (β=−.051; *P*=.02). Living for longer periods in a lockdown environment resulted in less emotional support via social media use (β=−.058; *P*=.03). More knowledge of deaths due to COVID-19 among family members and friends resulted in less informational support (β=−.064; *P*=.05) and emotional support (β=−.051; *P*=.03) via using social media. Both informational support (β=.165; *P*<.001) and peer support (β=.332; *P*<.001) were found to be associated with higher reported levels of VT. More peer support also increased levels of STS (β=.130; *P*<.001), whereas more emotional support led to an increase in VT levels (β=.144; *P*<.001), but not STS levels (*P*=.36).

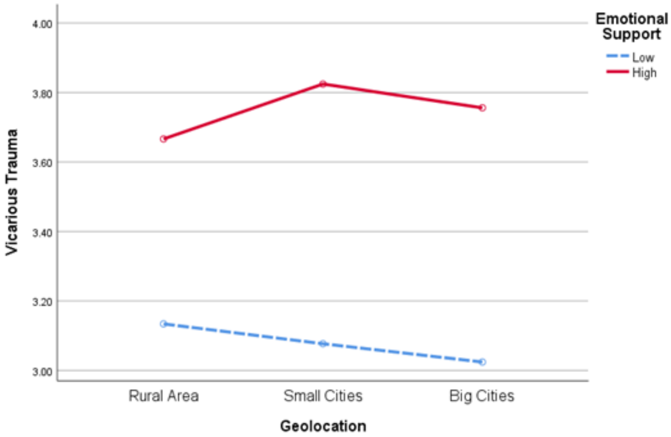
Figure 1. Structural equation modeling to predict secondary traumatic stress and vicarious trauma among participants during the COVID-19 outbreak in China.



Moderating Role of Geolocation

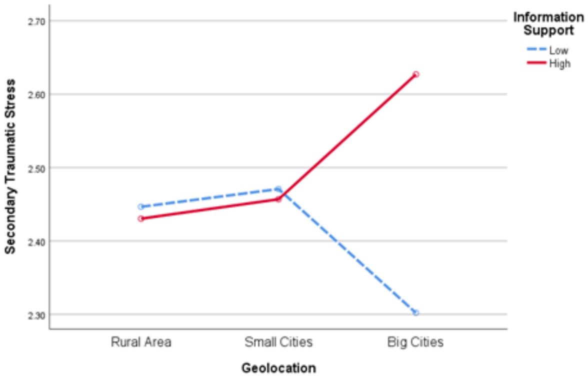
Geolocation had an interaction with emotional support in predicting VT (F_2=3.549; *P*=.029; see Figure 2). Participants from rural areas reported higher levels of VT when they received higher emotional support via social media (mean 3.666, SD 0.796) than when they received lower emotional support (mean 3.134, SD 0.606; t_{462}=7.947; *P*<.001); however, the differences were even larger for participants from small cities or towns (high emotional support: mean 3.825, SD 0.761 vs low emotional support: mean 3.077, SD 0.628; t_{913}=16.012; *P*<.001) and those from big cities (high emotional support: mean 3.756, SD 0.659 vs low emotional support: mean 3.024, SD 0.764; t_{307}=8.713; *P*<.001).

Figure 2. Interaction between geolocation and emotional support for vicarious trauma.



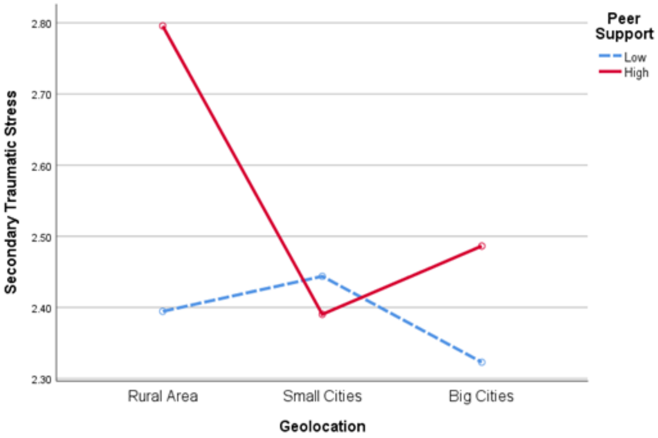
As shown in Figure 3, geolocation interacted with information support in predicting STS (F_2=5.093; P=.006). For participants who lived in rural areas or small cities, no significant difference in STS levels was found between those who received more information support and those who received relatively less information support via social media. However, among the participants who lived in big cities, those who received more information support via social media reported higher levels of STS (mean 2.627, SD 0.953) than those who received relatively less information support on social media (mean 2.302, SD 0.802; t_{299}=3.210; P=.001).

Figure 3. Interaction between geolocation and information support for secondary traumatic stress.



As shown in Figure 4, geolocation also moderated the relationship between peer support and STS (F_2=5.059; P=.006). For participants who lived in rural areas, those receiving more peer support experienced higher levels of STS (mean 2.796, SD 1.029) than those receiving less peer support (mean 2.395, SD 0.732; t_{350}=3.369; P=.001). However, for participants who lived in big and small cities, peer support on social media did not make a difference to the level of STS.

Figure 4. Interaction between geolocation and peer support for secondary traumatic stress.



Discussion

Principal Findings

As the ongoing COVID-19 pandemic affects people worldwide, it is crucial for researchers to conduct investigations that address the mental health consequences of the pandemic in order to mitigate the invisible harms caused to the general population. The findings of this study suggest that the mental health effects of the COVID-19 pandemic could have complex and multifactorial causes, including biological, behavioral, and environmental determinants such as social media usage. The analyses show that COVID-19 had taken a severe toll on the mental health of nonpatients, almost as soon as it started to spread in China. Social media use in addition to the lockdown environment and quarantine conditions have contributed the most to the overall toll on mental health.

Moreover, the analyses suggest that severe psychiatric disorders emerged in the general population of China, with 20% of the participants reporting anxiety, including 1 in 10 participants reporting moderate-to-severe anxiety. Furthermore, approximately 20% of the national sample reported depressive symptoms, of which 5.5% had moderate-to-severe depression. Overall, moderate levels of STS and considerably higher levels of VT were reported by these Chinese participants. The prevalence of psychiatric disorders reported seems particularly detrimental when we consider the facts that none of the participants were diagnosed with COVID-19 at the time of participation or had history of any psychiatric disorders before the COVID-19 outbreak.

Our findings suggest that a number of internal and external factors are related to stress, anxiety, and depression. Among the internal factors, participants who were younger, male, and more educated were more stressful or anxious than other participants at the peak of the COVID-19 outbreak in China. This finding is not consistent with previous studies in a nonpandemic context, which report that female participants tended to be more stressful. A possible explanation may be that younger individuals and men tended to go out more often to help others buy groceries than did older individuals or women, who spend more time at home in a lockdown situation [20]. Another explanation may be that compared to women, men are more susceptible to be stressed because of health reasons [51]. Previous studies have also shown that younger adults had a higher rate of psychological distress than older adults [54]. Future research should also examine different stress types and resilience mechanisms in relation to gender and age to understand how different stressors affect mental health differently.

The finding that social media use contributes to psychiatric disorders could be attributed to the fact that younger people relied more on social media and received more information about the pandemic through social media. It is worth noting that participants with higher income reported being more stressful and anxious than those who were less affluent. Younger, married, and more educated participants also reported having more STS symptoms, with no gender difference observed, suggesting that STS affected both men and women similarly. However, younger and more educated female participants reported higher levels of VT, which requires further research on why VT could affect women more than men.

Among the external factors, geolocation was found to be a critical determinant of mental health in COVID-19 nonpatients. In general, participants living in big cities felt similar levels of anxiety and depression as those living in small cities or towns, whereas participants in rural areas were the least stressful, anxious, and depressive. This may be due to the living conditions of big cities, which are more densely populated, and participants thus saw more infected patients and knew of more COVID-19–related deaths. In China, the medical resources are more readily available in cities than in rural areas; however, availability of more resources did not lead to less anxiety or fear among urban residents. This finding indicates that people in China had accurately assessed the severity of the emerging COVID-19 epidemic in the initial weeks.

Our findings suggest that even the individuals who were not themselves infected with COVID-19 nor were in quarantine could have experienced stress, anxiety, or depression as long as they lived in a lockdown situation or witnessed cases of infection, quarantine, and deaths among their family members or friends. The COVID-19 pandemic did not only endanger those infected with the virus but had also pushed China’s general population into a mental health crisis, and many people even reported that they experienced emotional trauma. All these occurrences were reported within 2 months of the COVID-19 outbreak. Thus, it is evident that living in the pandemic environment can be very debilitating and, in many cases, it has devastating effects on people’s psychological well-being with potentially lifelong consequences.

Another important finding of our study is the relationship between social media use for accessing health information and susceptibility of psychiatric disorders. This finding is consistent with those from a recent study in Wuhan, China—the first epicenter of COVID-19 globally, wherein researchers revealed that excessive social media use may lead to mental health issues [2]. In the present study, participants reported receiving social support from the health information shared on social media; this was especially true for those with more education and higher income. However, some differences were observed in terms of the specific informational, emotional, or peer support they received. Younger, female participants received the most informational support, whereas older participants received more emotional support than others, and participants who were married received more peer support than others. These results suggest that people process health information on social media differently and receive different types of support.

It is important to note that the approach does not suggest that social media use caused psychiatric disorders. We believe that an in-depth knowledge of social media use may contribute to a better understanding of the mechanisms of mental health conditions in a pandemic context.

On the other hand, knowledge of more quarantine cases during the outbreak was found to disrupt all the 3 types of support. Living longer in a lockdown environment had a similar effect of decreasing emotional support from social media use.

Knowledge of more cases of deaths among family members or friends led to them reporting less informational and emotional support. Meanwhile, participants who gained more informational and peer support reported lower levels of STS and VT. When people received more emotional support by using social medial, they tended to report higher levels of VT, but not STS. These findings call for more research on the pathological effects of STS and VT on the nonpatients in a pandemic environment.

It is noteworthy that although participants who knew of more quarantine cases reported receiving less informational, emotional, and peer support via social media, they still reported higher levels of STS and were more affected by mental illness. A possible explanation for this could be that participants who knew of more quarantine cases may have tried to seek information from other sources such as family, colleagues, and friends instead of via social media, and information from these sources may have increased their STS levels. Moreover, our additional data analyses showed that, for participants who knew of fewer than 4 quarantine cases, the information support (r=.069; P=.001), emotional support (r=.061; P=.005), and peer support (r=.113; P<.001) they received on WeChat were positively correlated to their STS levels. However, for participants who knew 4 or more quarantine cases, information support (P=.464), emotional support (P=.805), and peer support (P=.576) were no longer related to STS levels, suggesting that this group of participants generally maintained a high level of STS regardless of social media use. This finding is consistent with previous studies that described the potential anhedonia symptoms people experienced during COVID-19, which is characterized by the failure of experiencing pleasure from activities and is associated with depression, suicide, and other mental health issues [54].

Conclusion

The results of this study suggest that the general public is extremely vulnerable to mental health issues during the COVID-19 pandemic. Living in a pandemic situation can have serious mental health consequences even for those without any history of psychiatric disorders or not being infected with COVID-19 themselves. A range of internal and external factors have been identified that likely contribute to these mental health conditions. For instance, age, gender, marital status, education, and income levels play an important role in individuals' mental health conditions during the pandemic. Geolocation, lockdown duration, and social media use are also found to have an effect on mental health and traumatic disorders. Although people received health information support by using social media, excessive use of social media was found to be linked with elevated stress levels or psychiatric disorders. This finding does not suggest that social media use caused mental health issues but that it can mediate the levels of traumatic emotions experienced by people in the health crisis.

As the world is battling the COVID-19 pandemic, its detrimental effects on mental health will be more evident in the coming months and even after the pandemic is over. Health care providers thus need to carefully monitor psychosocial needs of the public and provide timely psychosocial support whenever needed. We believe that the findings in this research can help global citizens and health policy makers to mitigate psychiatric disorders in this as well as other public health crises, which should henceforth be regarded as a key component of general pandemic response.

Authors' Contributions

Bu Zhong and Zhibin Jiang share the co-first authorship.

Conflicts of Interest

None declared.

References

1. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A. The outbreak of COVID-19 coronavirus and its impact on global mental health. Int J Soc Psychiatry 2020 Jun;66(4):317-320. [doi: 10.1177/0020764020915212] [Medline: 32233719]
2. Zhong B, Huang Y, Liu Q. Mental health toll from the coronavirus: Social media usage reveals Wuhan residents' depression and secondary trauma in the COVID-19 outbreak. Comput Human Behav 2021 Jan;114:106524 [FREE Full text] [doi: 10.1016/j.chb.2020.106524] [Medline: 32836728]
3. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
4. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. Brain Behav Immun 2020 Oct;89:531-542 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
5. Bo H, Li W, Yang Y, Wang Y, Zhang Q, Cheung T, et al. Posttraumatic stress symptoms and attitude toward crisis mental health services among clinically stable patients with COVID-19 in China. Psychol Med 2020 Mar 27:1-2. [doi: 10.1017/s0033291720000999]
6. Zhang J, Lu H, Zeng H, Zhang S, Du Q, Jiang T, et al. The differential psychological distress of populations affected by the COVID-19 pandemic. Brain Behav Immun 2020 Jul;87:49-50 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.031] [Medline: 32304883]

7. Zhu S, Wu Y, Zhu C, Hong W, Yu Z, Chen Z, et al. The immediate mental health impacts of the COVID-19 pandemic among people with or without quarantine managements. Brain Behav Immun 2020 Jul;87:56-58 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.045] [Medline: 32315758]
8. Kang L, Ma S, Chen M, Yang J, Wang Y, Li R, et al. Impact on mental health and perceptions of psychological care among medical and nursing staff in Wuhan during the 2019 novel coronavirus disease outbreak: A cross-sectional study. Brain Behav Immun 2020 Jul;87:11-17 [FREE Full text] [doi: 10.1016/j.bbi.2020.03.028] [Medline: 32240764]
9. Smith C. The structural vulnerability of healthcare workers during COVID-19: Observations on the social context of risk and the equitable distribution of resources. Soc Sci Med 2020 Aug;258:113119 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113119] [Medline: 32534301]
10. Xie W, Newhagen JE. The effects of communication interface proximity on user anxiety for crime alerts received on desktop, laptop, and hand-held devices. Communic Res 2012 Jun 05;41(3):375-403. [doi: 10.1177/0093650212448670]
11. Lau S. How China’s industrial internet is fighting COVID-19. World Economic Forum Internet. 2020 Apr 08. URL: https://www.weforum.org/agenda/2020/04/china-covid-19-digital-response [accessed 2020-12-21]
12. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
13. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
14. Marino C, Gini G, Vieno A, Spada MM. The associations between problematic Facebook use, psychological distress and well-being among adolescents and young adults: A systematic review and meta-analysis. J Affect Disord 2018 Jan 15;226:274-281. [doi: 10.1016/j.jad.2017.10.007] [Medline: 29024900]
15. Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, et al. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. PLoS ONE 2017 Aug 16;12(8):e0181722. [doi: 10.1371/journal.pone.0181722]
16. Benuto L, Singer J, Cummings C, Ahrendt A. The Vicarious Trauma Scale: Confirmatory factor analysis and psychometric properties with a sample of victim advocates. Health Soc Care Community 2018 Jul;26(4):564-571. [doi: 10.1111/hsc.12554] [Medline: 29488272]
17. Li X, Liu Q. Social media use, eHealth literacy, disease knowledge, and preventive behaviors in the COVID-19 pandemic: Cross-sectional study on Chinese netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
18. Katz E, Haas H, Gurevitch M. On the use of the mass media for important things. American Sociological Review 1973 Apr;38(2):164. [doi: 10.2307/2094393]
19. Katz E, Blumler JG, Gurevitch M. Utilization of Mass Communication by the Individual. In: Blumler JG, Katz E, editors. The Uses of Mass Communications: Current Perspectives on Gratifications Research. Beverly Hills, CA: Sage Publications; 1974:19-32.
20. Zhao J, Han H, Zhong B, Xie W, Chen Y, Zhi M. Health information on social media helps mitigate Crohn's disease symptoms and improves patients' clinical course. Comput Human Behav 2021 Feb;115:106588. [doi: 10.1016/j.chb.2020.106588]
21. Stephens MH, Petrie KJ. Social Support and Recovery from Disease and Medical Procedures. In: Wright JD, editor. International Encyclopedia of the Social & Behavioral Sciences (Second Edition). Amsterdam: Elsevier; 2015:735-740.
22. Gudykunst W, editor. Theorizing about intercultural communication. Thousand Oaks, CA: Sage Publications; 2005:480.
23. Helgeson VS. Two important distinctions in social support: Kind of support and perceived versus received. J Appl Social Pyschol 1993 May;23(10):825-845. [doi: 10.1111/j.1559-1816.1993.tb01008.x]
24. Ko H, Wang L, Xu Y. Understanding the different types of social support offered by audience to A-list diary-like and informative bloggers. Cyberpsychol Behav Soc Netw 2013 Mar;16(3):194-199 [FREE Full text] [doi: 10.1089/cyber.2012.0297] [Medline: 23363225]
25. Mead S, Hilton D, Curtis L. Peer support: a theoretical perspective. Psychiatr Rehabil J 2001;25(2):134-141. [doi: 10.1037/h0095032] [Medline: 11769979]
26. Langford CPH, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. J Adv Nurs 1997 Jan;25(1):95-100. [doi: 10.1046/j.1365-2648.1997.1997025095.x] [Medline: 9004016]
27. Shim M, Jo HS. What quality factors matter in enhancing the perceived benefits of online health information sites? Application of the updated DeLone and McLean Information Systems Success Model. Int J Med Inform 2020 May;137:104093. [doi: 10.1016/j.ijmedinf.2020.104093] [Medline: 32078918]
28. Hibbard JH. Patient activation and the use of information to support informed health decisions. Patient Educ Couns 2017 Jan;100(1):5-7. [doi: 10.1016/j.pec.2016.07.006] [Medline: 27432014]
29. McKenna KYA, Bargh JA. Coming out in the age of the Internet: Identity "demarginalization" through virtual group participation. J Pers Soc Psychol 1998;75(3):681-694. [doi: 10.1037/0022-3514.75.3.681]

30. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
31. Mehnert A, Lehmann C, Graefen M, Huland H, Koch U. Depression, anxiety, post-traumatic stress disorder and health-related quality of life and its association with social support in ambulatory prostate cancer patients. Eur J Cancer Care (Engl) 2010 Nov;19(6):736-745. [doi: 10.1111/j.1365-2354.2009.01117.x] [Medline: 19832893]
32. Rogala L, Miller N, Graff LA, Rawsthorne P, Clara I, Walker JR, et al. Population-based controlled study of social support, self-perceived stress, activity and work issues, and access to health care in inflammatory bowel disease. Inflamm Bowel Dis 2008 Apr;14(4):526-535. [doi: 10.1002/ibd.20353] [Medline: 18183608]
33. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
34. Yang F, Zhong B, Kumar A, Chow S, Ouyang A. Exchanging social support online: a longitudinal social network analysis of irritable bowel syndrome patients' interactions on a health forum. Journal Mass Commun Q 2017 Sep 28;95(4):1033-1057. [doi: 10.1177/1077699017729815]
35. Dai E, Sun Y, Wang S. Ginger cannot cure cancer: battling fake health news with a comprehensive data repository. Palo Alto, CA: 2020; 2020 May 26 Presented at: Proceedings of the Fourteenth International AAAI Conference on Web and Social Media; 2020; Atlanta, GA p. 1-11 URL: https://arxiv.org/abs/2002.00837 [doi: 10.5281/zenodo.3606757]
36. Hagerty SL, Williams LM. The impact of COVID-19 on mental health: The interactive roles of brain biotypes and human connection. Brain Behav Immun Health 2020 May;5:100078 [FREE Full text] [doi: 10.1016/j.bbih.2020.100078] [Medline: 32382727]
37. Keshavan MS. Pandemics and psychiatry: Repositioning research in context of COVID-19. Asian J Psychiatr 2020 Jun;51:102159 [FREE Full text] [doi: 10.1016/j.ajp.2020.102159] [Medline: 32442894]
38. Hao K, Basu T. The coronavirus is the first true social-media "infodemic". MIT Technology Review Internet. 2020 Feb 12. URL: https://www.technologyreview.com/2020/02/12/844851/the-coronavirus-is-the-first-true-social-media-infodemic/ [accessed 2020-12-21]
39. Wiederhold BK. Using social media to our advantage: alleviating anxiety during a pandemic. Cyberpsychol Behav Soc Netw 2020 Apr;23(4):197-198. [doi: 10.1089/cyber.2020.29180.bkw] [Medline: 32207630]
40. Brummett BH, Babyak MA, Siegler IC, Vitaliano PP, Ballard EL, Gwyther LP, et al. Associations among perceptions of social support, negative affect, and quality of sleep in caregivers and noncaregivers. Health Psychol 2006 Mar;25(2):220-225. [doi: 10.1037/0278-6133.25.2.220] [Medline: 16569114]
41. Paykel ES. Historical overview of outcome of depression. Br J Psychiatry 2018 Aug 06;165(S26):6-8. [doi: 10.1192/s0007125000293227]
42. Michalak EE, Wilkinson C, Hood K, Dowrick C, Wilkinson G. Seasonality, negative life events and social support in a community sample. Br J Psychiatry 2018 Jan 02;182(5):434-438. [doi: 10.1192/bjp.182.5.434]
43. Mohr DC, Classen C, Barrera M. The relationship between social support, depression and treatment for depression in people with multiple sclerosis. Psychol Med 2004 Apr;34(3):533-541. [doi: 10.1017/s0033291703001235] [Medline: 15259838]
44. Revenson TA, Schiaffino KM, Deborah Majerovitz S, Gibofsky A. Social support as a double-edged sword: The relation of positive and problematic support to depression among rheumatoid arthritis patients. Social Science & Medicine 1991 Jan;33(7):807-813. [doi: 10.1016/0277-9536(91)90385-p]
45. Andreassen CS, Pallesen S, Griffiths MD. The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. Addict Behav 2017 Jan;64:287-293. [doi: 10.1016/j.addbeh.2016.03.006] [Medline: 27072491]
46. Oh HJ, Lee B. The effect of computer-mediated social support in online communities on patient empowerment and doctor-patient communication. Health Commun 2012;27(1):30-41. [doi: 10.1080/10410236.2011.567449] [Medline: 21797714]
47. Barrera MJ, Glasgow R, McKay H, Boles S, Feil E. Do internet-based support interventions change perceptions of social support?: An experimental trial of approaches for supporting diabetes self-management. Am J Community Psychol 2002 Oct;30(5):637-654. [doi: 10.1023/A:1016369114780] [Medline: 12188054]
48. Lovibond P, Lovibond S. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-U]
49. Cieslak R, Shoji K, Douglas A, Melville E, Luszczynska A, Benight CC. A meta-analysis of the relationship between job burnout and secondary traumatic stress among workers with indirect exposure to trauma. Psychol Serv 2014 Feb;11(1):75-86. [doi: 10.1037/a0033798] [Medline: 23937082]
50. Bride BE, Robinson MM, Yegidis B, Figley CR. Development and validation of the Secondary Traumatic Stress Scale. Research on Social Work Practice 2016 Aug 19;14(1):27-35. [doi: 10.1177/1049731503254106]
51. Pearlman L, Saakvitne K. Trauma and the Therapist: Countertransference and Vicarious Traumatization in Psychotherapy with Incest Survivors. New York, NY: W.W. Norton; 1995:451.

52. Vrklevski LP, Franklin J. Vicarious trauma: the impact on solicitors of exposure to traumatic material. Traumatology 2008;14(1):106-118. [doi: 10.1177/1534765607309961]
53. Population of Provinces, Municipalities, and Autonomous Regions, 2010 Census of China, and Change in the Percent Distribution by Area, 2000-2010. Population Reference Bureau. Beijing, China: Population Reference Bureau; 2011. URL: https://www.prb.org/wp-content/uploads/2011/05/china-2010-census-results-table.pdf [accessed 2020-12-21]
54. Twenge JM, Cooper AB, Joiner TE, Duffy ME, Binau SG. Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005-2017. J Abnorm Psychol 2019 Apr;128(3):185-199. [doi: 10.1037/abn0000410] [Medline: 30869927]

Abbreviations

DASS-21: 21-item Depression Anxiety Stress Scale
RQ: research question
STS: secondary traumatic stress
VT: vicarious trauma

Edited by G Fagherazzi; submitted 20.08.20; peer-reviewed by A Ramazan, L Iglesias; comments to author 26.09.20; revised version received 23.10.20; accepted 09.12.20; published 31.12.20.

Please cite as:
Zhong B, Jiang Z, Xie W, Qin X
Association of Social Media Use With Mental Health Conditions of Nonpatients During the COVID-19 Outbreak: Insights from a National Survey Study
J Med Internet Res 2020;22(12):e23696
URL: https://www.jmir.org/2020/12/e23696
doi:10.2196/23696
PMID:33302256

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org